International Conference Workshop on Rhythm In Speech and Music From Neruro-Cognitive Perspectives

Guest Editors

Dipak Ghosh

Shankha Sanyal, Pijush Kanti Gayen, Ratul Ghosh

A JJLL Publication of School of Languages and Linguistics
Jadavpur University
Kolkata
2020

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics

ISSN: 2581-494X

Volume 4, Special Issue

Mean Length of Utterance of Word and Morpheme for 3-5 years Bengali speaking children 01-13
by Gajendra Gouda, Suman Kumar, Mita Sarkar, Richa Rashmi, Nikita Chatterjee, and Susmi Pani
Unintentional Voice Modulations in Real World Conversations and its impact in Automatic Speaker 14-25
Recognition by Soma Khan, Jayanta Basu, Rajib Roy, Madhab Pal and Milton S. Bepari
Timbre Based Style Identification 26-32
by Kaushik Banerjee, Anirban Patranabis, Ranjan Sengupta and Dipak Ghosh
The cognitive aspect of word meaning in the Brahmakanda of Vakyapadiyam 33-39
by Rusa Bhowmik
An Acoustical and Neuro-cognitive Study on the Effects of Lyrics in Song from Non-Linear 40-53
Perspective by Archi Banerjee, Shankha Sanyal, Souparno Roy, Priyadarshi Patnaik, & Dipak Ghosh
Speech Rhythm in Malayalam Speaking Children with Hearing Impairment 54-60
by Yeshoda Krishna, Revathi Raveendrum, & Sreeraj Konadath
Meaning Making through 'Music' and 'Emotion' in Bengali Children's Rhymes (Choras) 61-68
by Rajoshree Chatterjee & Jayshree Chakraborty
How does musical notes correlate with human emotion? A psycho-acoustic exploration with Indian 69-81
Classical Music by Medha Basu, Archi Banerjee, Shankha Sanyal, & Dipak Ghosh
Finding Geometry in Quantifier\, A Cognitive Perspective 82-97
by Spandan Chowdhury
In Search of Rhythm: A Correlational Study between different Emotive Poetry pieces and their 98-106
corresponding preliminary Emotional Categorization using Fractal Analytics – A Pilot Study
by Ratul Ghosh, Shankha Sanyal, Samir Karmakar, & Dipak Ghosh
Development and standardisation of Bengali sentence identification test 107-120
by Mousumi Chatterjee, Indranil Chatterjee, Palash Dutta, & Krishna Kali Banerjee
Communicative Form vs. Literary Form: An intervention inspired by Nigel Fabb 121-129
by Rimi Ghosh Dastidar
Improvisation in Indian Classical Music: Probing with M-B and B-E distributions 130-143
by Souparno Ray, Archi Banerjee & Shankha Sanyal
Ornamentation in Hindustani Music 144-150
by Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta & Dipak Ghosh
A InTraSAL (Intonational Transcription of South Asian Languages) analysis of Standard Colloquial 151-162
Bengali by Moumita Pakrashi & Shakuntala Mahanta
Cognitive Functions in Non-musicians in Music Perception 163-174
by Sukdeb Goswami
Fractal Based Categorization of Bengali Phonemes: A Pilot Study 175-181
by Pijush Kanti Gayen, Shankha Sanyal & Samir Karmakar

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Mean Length of Utterance of Word and Morpheme for 3-5 years Bengali speaking children

Gajendra Gouda[1], Suman Kumar[2], Mita Sarkar[2], Richa Rashmi[2], Nikita Chatterjee[2], Susmi Pani[2]
[1]Starkey Hearing Technologies
[2]Department of Speech Language Pathology, AYJNISHD(D), RC, India

ARTICLE INFO

Article history:
Received 12/03/2020
Accepted 05/08/2020

Keywords:
MLU,
Bengali speaking children,
morpheme,
word,
irregularity,
acoustic analysis

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Background: Mean Length of Utterance (MLU) is a better predictor of language age than the chronological age and its usefulness is increasing day by day essentially in child language assessment. There have been no studies exploring the MLU in Bengali speaking children. This study examines the utility of the Mean Length of Utterance in morpheme (MLUm) and the Mean Length of Utterance in word (MLUw) in Bengali context.

Methods: The participants included in this study were 200 native Bengali speaking children of age range 3 to 5 years. The age range is divided into four equal age groups with similar male and female ratio. Utterances were elicited from each child using picture cards. Descriptive statistics, t-test and paired t-test was used for statistical analysis.

Results: The results obtained indicate that the MLUw and MLUm of the 3 years to 3 year 6 months male children were 1.8 and 2.8 similarly in female children 2.2 and 3.4 respectively. The MLUw and MLUm obtained from the children of 4 year 6 months to 5 years male children were 3.3 and 4.1 similarly in female children, 3.7 and 4.9 respectively.

Conclusion: The study concluded that there was a significant difference between the acquisition of MLUm and the MLUw in Bengali speaking children. The MLUm and the MLUw were acquired more in number by the female children than the male children and the MLUs are increasing with increase in age. The MLUs obtained in morphemes and words for the age range 3 years to 5 years Bengali speaking children may be used for clinical purpose, clinical research and determining prognosis as it can serve the purpose of normative.

1. INTRODUCTION

An utterance is a vocal expression preceded and followed by silence; may be made up of words, phrases, clauses or sentences. A word is a free form consisting of a sequence of one or more phonemes and one or more syllables which have meaning without being divisible into smaller units capable of independent use; e.g., I, go, mother, baby etc. A morpheme is the smallest meaningful unit of language having a differential function which is of various types; e.g., bound morpheme like -ed, -ing, etc, free morpheme

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***
Corresponding Author: Richa Rashmi
Email: richa.avasthi94@gmail.com

like cat, walk, soft etc. Roger Brown in 1973[1] proposed a comprehensive method of calculating the length of utterance is called the Mean Length of Utterance in morphemes (MLUm). MLUm is computed by dividing the total number of morphemes in an utterance by the total number of utterances. Brown (1973) proposed that

MLUm as 'an excellent simple index of grammatical development'. Brown (1973) divided the grammatical development of a children based on the MLUm into five stages;

Stage I: Semantic roles and syntactic relations (MLU 1.0 - 2.0 morphemes or 1.75 morphemes). Here child puts noun-verb sequences together.

Stage II: Grammatical morphemes and modulation meaning (MLU = 2.0 - 2.5 or 2.25 morphemes). The child starts to change word endings to portray grammar.

Stage III: Modalities of simple sentences (MLU = 2.5 - 3.25 or 2.75morphemes). The child begins to use questions and imperatives.

Stage IV: Embedding (MLU = 3.25 - 3.75 or 3.5 morphemes). The child begins to Use complex sentences.

Stage V: Co-ordination (MLU = 3.75 - 4.25 or 4 morphemes). The child may use connectors and more functions.

The average amount of words or morphemes that a person produces in each utterance is called as the mean length of utterance. The Mean Length of Utterance (MLU) provides useful information about language development and it is an important indicator of language disorder or delay. Generally, a normal child's chronological age (up to age 5) will correspond closely to his or her MLU. For example, a normally developing 4 year, 3-month-old child will often exhibit a MLU of approximately 4.3 plus or minus a few tenths. The Mean Length Utterance (MLU) is one of the language measurements that can be obtained through spontaneous discourse. Its main goal is obtaining data about morphological and syntactical aspects of language in children with both typical development and with language disorders (Miller & Chapman, 1981)[2].

Use of MLU

Several qualitative and quantitative procedures are adopted in attempts to describe and assess the language of children. One such procedure which is found to be particularly useful with the clinical population of developmentally disabled children is computing mean length of utterance in words/morphemes. It provides an index of syntactic complexity in the child's speech. The Mean Length of Utterance (MLU) has gained sustained popularity and interest of the professionals for long, for its relative ease of use and precision. It successfully serves as a tool for identifying language delay and deviancy. It is a more accurate measure of language acquisition than chronological age of a child. MLU is increasingly used with language disordered population as it serves as a tool for identifying language delay and deviances (Dessai, 2009)[3].

MLU and Language Age of Children

Mean Length of Utterance becomes helpful to find out the language age of children who is having developmental language disorder. According to Parker (2005)[4] who compared MLU (w) and MLU (m) scores of 40 language transcripts from typically developing English speaking children between the ages of 3:0 and 3:10. Results indicated that MLU (m) and MLU (w) are almost perfectly correlated. This finding suggests that MLU (w) can be used as effectively as MLU (m) as a measurement of a child's gross language development.

1. AIM OF THE STUDY

To develop a normative for Mean Length of Utterance in morpheme (MLUm) and Mean Length of Utterance in word (MLUw) among Bengali speaking children (3-5 years) to assess the language development in these children.

2. OBJECTIVE

- To obtain the Mean Length of Utterance at the level of morpheme and word for both male & female children between all the age groups.
- To compare between acquisition of the MLUm and MLUw between genders and in between different age groups.

3. NEED OF THE STUDY

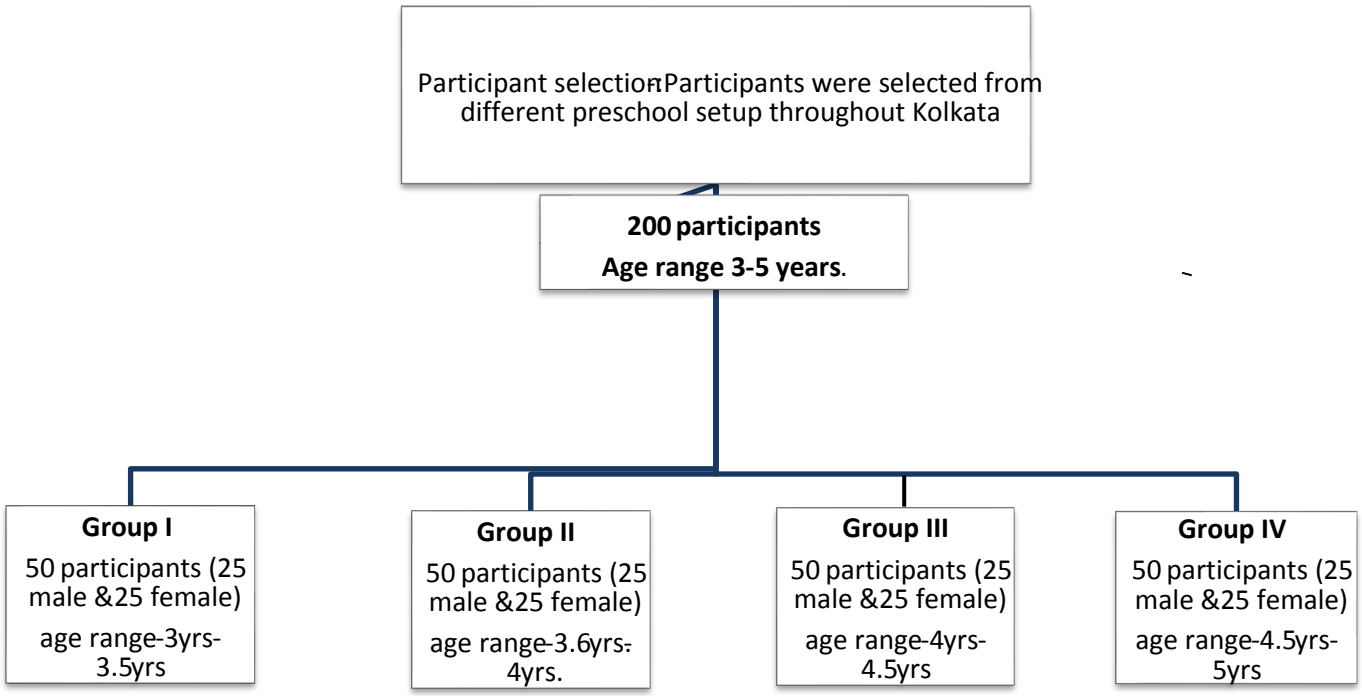
It is important to measure the language development among children to evaluate the efficacy of language stimulation. MLU is one of the major indicators for language development. However, there are very few literatures exists on mean length of utterance in Indian languages and dearth of literature on the normative of Bangla language to assess language development. Therefore, the aim of the current study is to develop the normative for mean length of utterance for the age range of 3-5 years. The expected results will be helpful for making a decision on how much the children lack in terms of language than their typically developing peers.

STUDY DESIGN

Cross-Sectional Study design.

4. METHODOLOGY

Participants: 200 participants would be taken within the age range of 3 to 5 years and the participants will be divided into four age ranges. Each age range will contain similar number of male and female participants.



Participant selection criteria

Inclusion Criteria

- All the participants are within the age range of 3 to 5 years.
- All the participants will have similar competence in receptive and expressive language.
- All the participants have normal hearing sensitivity.
- All the participants are native speakers of Bengali language.
- No history of speech, language, hearing, psychological difficulty.

5. Research Hypotheses

Hypothesis: There would be no significant difference in MLUs of male & female children across all the age groups for both morpheme and word level.

Material

- The questions of daily activities.

- Simple Bengali stories.Topic cards will be given such as the picnic, The park, The busy street, The market, Railway Station, The village, The zoo, The Birthday party

6. Procedure

The Communication DEALL will be performed to correlate the language competence equivalency of the participants. Then the desired materials will be used and the participant's 100 random utterances will be recorded. The mean length of utterance will be calculated by using following formula;

Mean Length of Utterance (Word)= Number of words/Total no of utterances
Mean Length of Utterance (Morpheme)= Number of morphemes/Total no. of utterances
Shipley (2009)

Recording of Speech Sample:

- The verbal responses obtained were recorded on a high-fidelity digital tape recorder during investigator-child interaction. Positive reinforcement was given whenever necessary.
- All the speech samples will be taken if intelligible and clearly understood.

7. RESULT

The use of Mean Length of Utterance (MLU) became an easy and essential tool to rule out the language delays in children after the study by Brown (1973). The present study is an attempt to provide the Mean Length of Utterance in word and Mean Length of Utterance in morpheme of Bengali speaking children of 35 years. The descriptive statistics was obtained comprising Mean, Standard deviation of the MLUm and MLUw in both male and female participants across the age groups to establish respective values of MLUs.

Age Range of Groups	Gender	Mean Age		Mean		Standard Deviation	
		MLUm	MLUw	MLUm	MLUw	MLUm	MLUw
Group I (3yrs-3.6yrs)	Male	3.2	3.2	2.81	1.8	0.1	0.15
	Female	3.2	3.2	3.47	2.24	0.15	0.17
Group II (3.6yrs-4yrs)	Male	3.6	3.6	3.22	2.2	0.19	0.08
	Female	3.6	3.6	3.63	2.62	0.21	0.13
Group III (4yrs-4.6yrs)	Male	4.1	4.1	3.8	2.61	0.11	0.12
	Female	4.1	4.3	4.29	3.26	0.14	0.23
Group IV (4.6yrs-5yrs)	Male	4.6	4.6	4.18	3.36	0.11	0.16
	Female	4.6	4.6	4.9	3.77	0.16	0.14

Table 1: The Mean and Standard Deviation (S.D) of Mean Length of Utterance in words & morphemes for both male & female children among all the four different ranges of age group.

Table 1 depicted that the mean of MLUw of the male participants of age group IV obtained the highest mean of 3.36 and the participants of age group I obtained the lowest mean of 1.86. The female children of age group IV obtained the highest mean 3.77 and the children of age group I obtained 2.24 as the lowest mean. Similarly, MLUm of male children of age group IV obtained the highest mean 4.18 and the children of age group I obtained 2.81 as the lowest mean. The female children of age group IV obtained the highest mean 4.9 and the children of age group I obtained 3.47 as the lowest mean.

Groups	Gender	T	df	P value
Group I	**Male**	**-30.386**	**24**	**0.000**
	Female	**-25.232**	**24**	**0.000**
Group II	**Male**	**-30.386**	**24**	**0.000**
	Female	**-21.553**	**24**	**0.000**
Group III	**Male**	**-32.129**	**24**	**0.000**
	Female	**-19.367**	**24**	**0.000**
Group IV	**Male**	**-20.435**	**24**	**0.000**
	Female	**-25.284**	**24**	**0.000**

Table 2: Comparison of acquisitions of MLUm versus MLUw in male children and in female children across all the age groups.

The table 2 represented the paired t-test results which indicated that there is significant difference in the MLUm and the MLUw acquisitions among male children and also in female participants of all the age groups with p value 0.000 at the 0.05 level of significance.

Groups	**MLU**	**T**	**Df**	**P value**
Group I	**MLUm**	**-17.880**	**48**	**0.000**
	MLU w	**-9.578**	**48**	**0.000**
Group II	**MLUm**	**-7.162**	**48**	**0.000**
	MLUw	**-13.228**	**40.767**	**0.000**
Group III	**MLUm**	**-13.025**	**48**	**0.000**
	MLUw	**-12.402**	**35.951**	**0.000**
Group IV	**MLUm**	**-18.405**	**48**	**0.000**
	MLUw	**-9.396**	**48**	**0.000**

Table 3: Comparison of Male versus Female children in acquisition of MLUm and also in MLUw across all the age groups.

The table 3 provided the results of the t-test which indicated that there was significant difference in both the male and female participants in acquiring the MLUm as well as MLUw among all the age groups with p value 0.000 at the 0.05 level of significance ($p < 0.05$).

Age Group Comparison	Gender	T		Df		P value		Mean Difference	
		MLm	MLUw	MLUm	MLUw	MLUm	MLUw	MLUm	MLUw
Group I&II	Male	9.464	11.516	48	48	.000	.000	.413	.401
	Female	3.052		48	48	.004	.000	.161	.376
Group II&III	Male	12.781	13.834	48	48	.000	.000	.578	.409
	Female	12.634		48	48	.000	.000	.660	.637
Group III&IV	Male	11.619	18.227	48	48	.000	.000	.378	.749
	Female	13.792		48	48	.000	.000	.604	.510

Table 4: t test results depicting mean differences in acquisition of MLUm & MLUw as obtained by
comparison between successive age groups in male and female children.

Table 4 depicts the t test results revealing mean differences that there was significant difference in acquisition of MLUm between male children of age group I &II ,II &III and III & IV with p value <0.05 and also, there was significant difference in acquisition of MLUm between female children of age group I & II,II&III,III&IV where p value < 0.05.
Similarly, there was significant difference in acquisition of MLUw between male children of age group I &II, II &III and III & IV with p value <0.05 and also, there was significant difference in acquisition of MLUw between female children of age group I&II,II&III,III&IV where p value < 0.05.Which indicate there is significant difference in comparison of successive age groups for male and female children.

8. DISCUSSION

The aim of the current study was to measure the Mean Length of Utterance in morpheme (MLUm) and the Mean Length of Utterance in word (MLUw) in the Bengali speaking children without any language impairment. This study reveals that MLUm and MLUw development in Bengali children increases with increasing in age and MLUm maintains to be higher in score than the MLUw throughout the age. Growth of the MLUm and the MLUw in both male and female children in the age range of 3- 5 years shown in the figures1 and 2 and it is evident that there is gradual increase in number of words and morpheme acquisitions, beginning at 3 years of age till the upper age limit of the study that is 5 years of age.

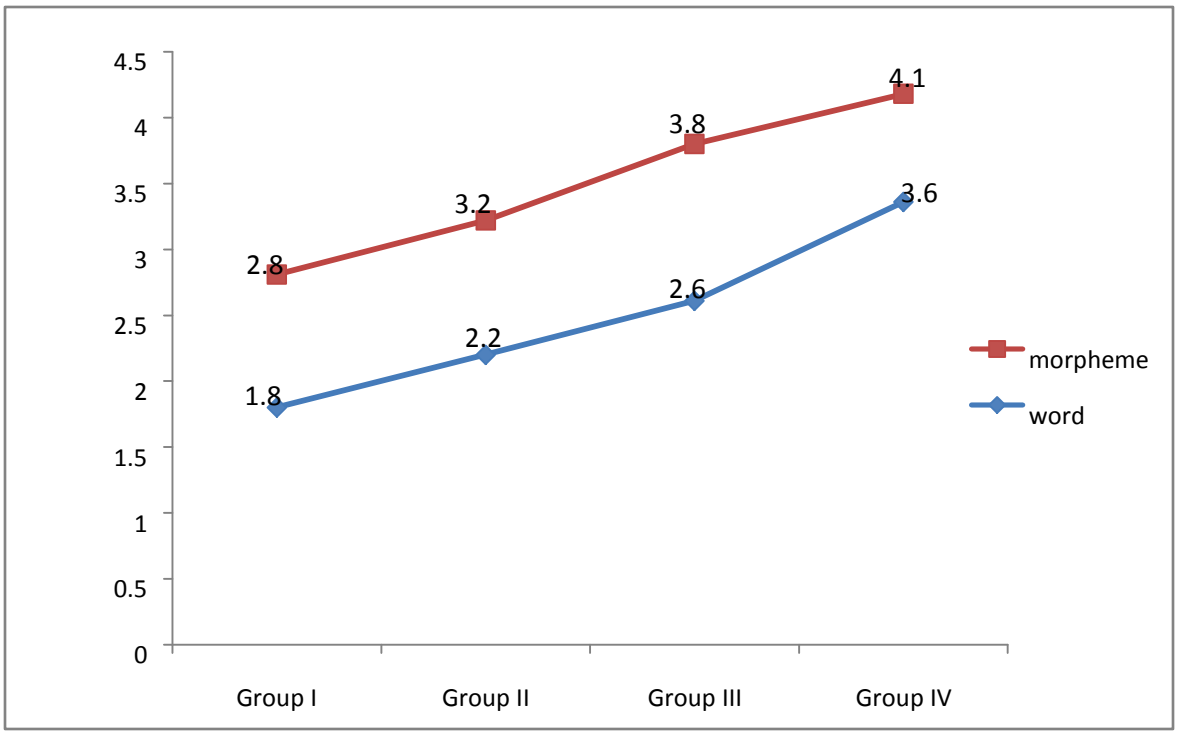


Figure 1: Acquisition of MLUm and the MLUw in male children across group.

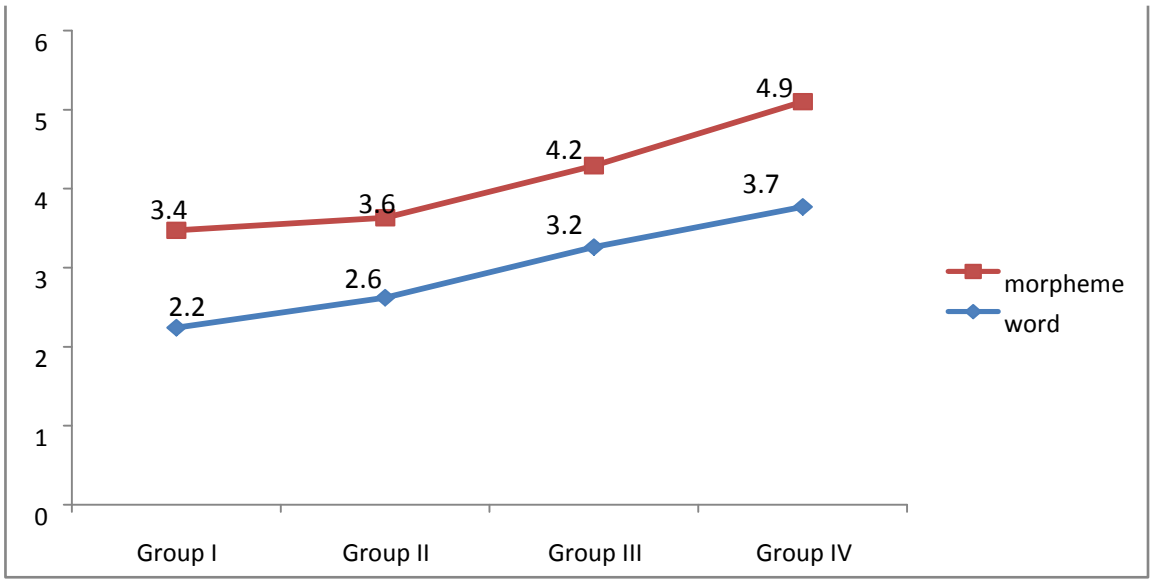


Figure 2 The acquisition of MLUm and MLUw in female children across group.

The relationship of MLUm and MLUw with increase in age suggests that the words and morphemes increase almost similarly in this age range 3 to 5 years but the morphemes are more in number than the words.

Similarly, the same results were obtained by the other researchers like Dessai and Karanth (2009) in a study on Konkani language found that the MLUm and MLUw increases with increase in age.

Miller and Chapman (1981) with 123 midwestern children (17-59 mts) without any significant language difficulty found that the MLU increases with increase in age and atypical increase in MLU usually seen after 5 years of age.

Hickey (1991)[5] in a study on Irish language also found that the MLUm and the MLUw increases significantly with increase in age and both of these can be used as effective measure of language development.

Contrastive findings reported by Parker and Brorson (2005) in a study on native English children and found that there was no significant difference between the MLUm and the MLUw.

Several researchers have found that the MLUm was significantly correlated with MLUw in different languages like Snow et al., (1978)[6] in Dutch, Hickey (1991) in Irish and Thordardottir and Weismer (1998)[7] in Icelandic language which have shown that the MLUw was an effective measure of language like MLUm.

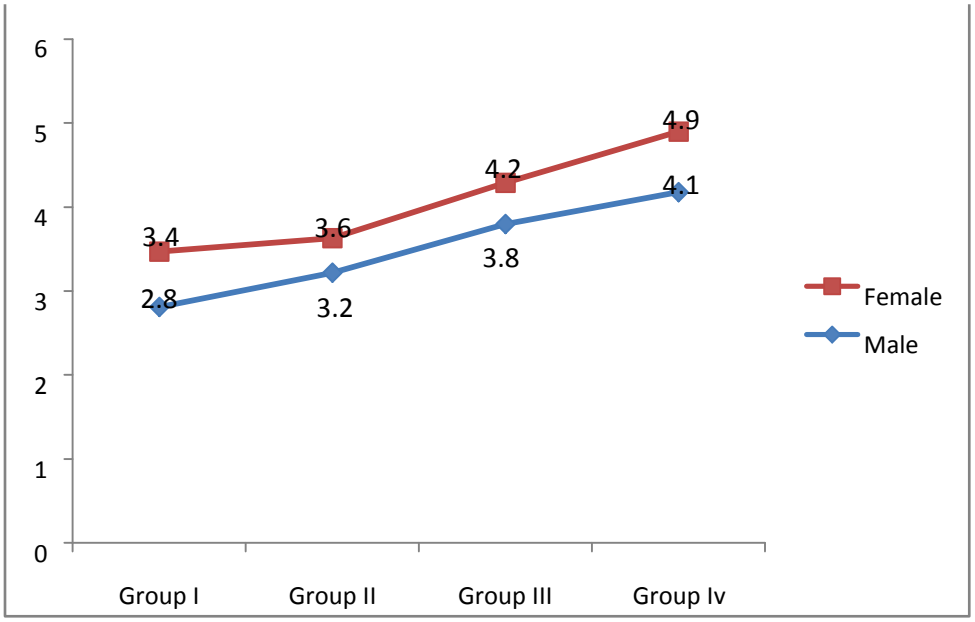


Figure 3: The acquisition of MLUm between male and female children across groups.

Table 3 depicts that the MLUm acquisition is significantly more in female children than male children in all the age groups.

These findings are in consonance with the findings of Schachter et al., (1978)[8] in which they also found that the MLU increases more in females than males.

Contrastive findings were also obtained by Hickman (2000)[9] in a study on the sex difference in 2-3-year children using Mean Length of Utterance and found that there was no significant difference in the MLU of male and female children. From the figure 3 it is evident that females acquire MLUm more in number in all the age groups which means that their acquisition of MLUm is always higher right from the beginning age that is 3 years till the upper age limit (5 years) of this study.

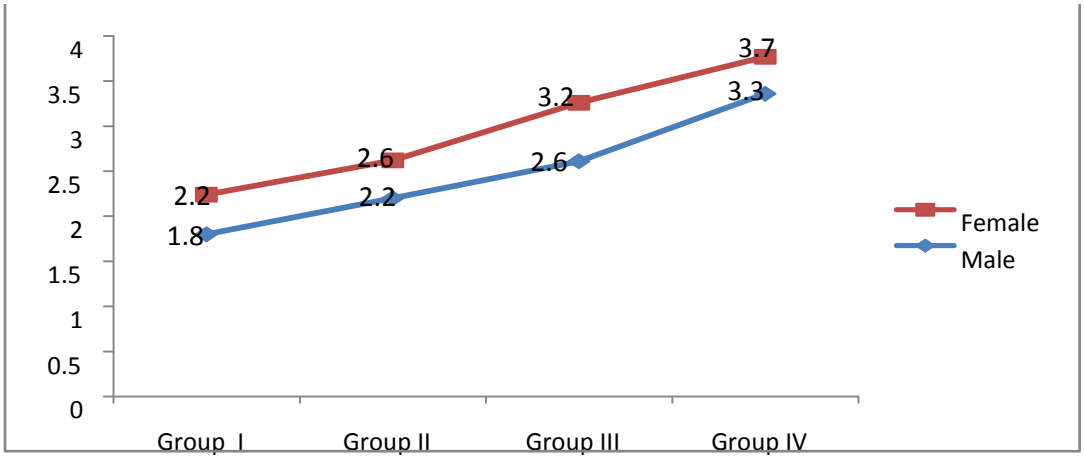


Figure 4: The acquisition of MLUw between male and female children across groups.

Table 3 depicts that the MLUw acquisition is significantly more in female children than male children in all the age groups.

Present findings are in consonance with the findings of Schachter et al. (1978) in a study on child language development specifically the language development of the girl children and found that younger toddler girls were significantly advanced in Mean Length of utterance of words and morphemes.

A contrastive finding was also obtained by Voniati (2016)[10]. They conducted a study on the MLU in Cypriot Greek speaking children and found that there was no significant difference of MLUw between male and female children. From the figure 4 it is evident that females acquire MLUw more in number in all the age groups which means that their acquisition of MLUw is always higher right from the beginning age that is 3 years till the upper age limit (5 years) of this study.

Therefore, the findings are indicating that there is a higher preponderance in acquiring morphemes and words in Bengali female children than male children in the age range of 3 to 5 years.

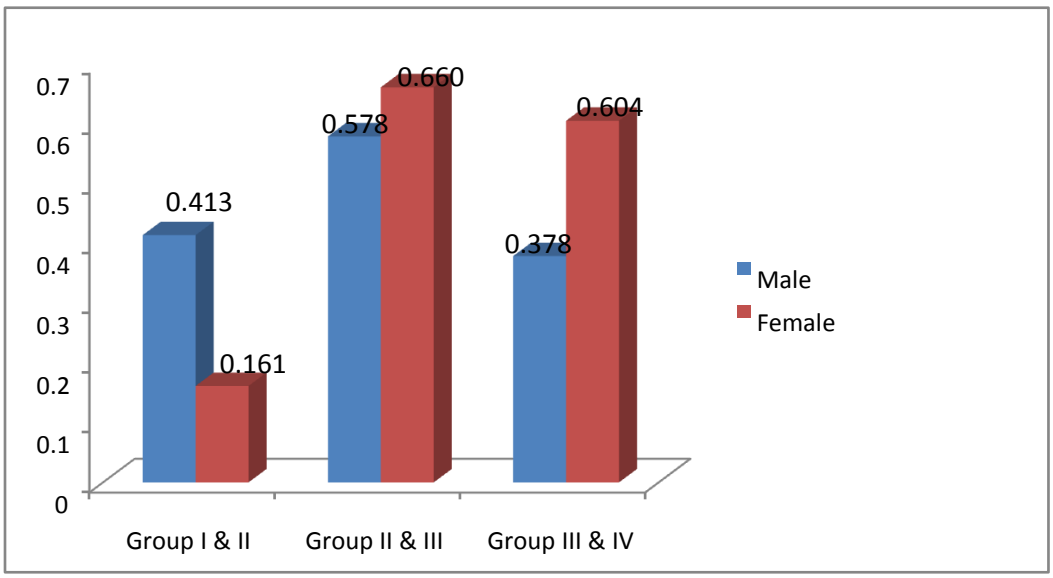


Figure 5: Comparison between group I & II, group II & III, group III & IV on the basis of MLUm acquisition male and female children.

In the tables 4 in the result section, the mean difference is 0.413 between age group I and II which indicates that male children of 3 years to 3 year 6 months are significantly lower in acquiring morphemes than the male children of 3 year 6 months to 4 years and similar findings was obtained in female children though the mean difference is much lower (0.161) than the male children.

The mean difference between the age group II and III which indicates that the male children of 3 year 6 months to 4 years are significantly lower in acquiring morphemes (mean difference is 0.578) than the male children of 4 years to 4 year 6 months and similar findings was obtained in female children with mean difference 0.660.

The mean difference between the age group III and IV also indicates that there is significantly lower number of morphemes acquired in male children in the age range of 4 years to 4 year 6 months than 4 year 6 months to 5 years with mean difference 0.378 and similarly significant difference was obtained in female children with mean difference 0.604.

These findings are depicted in figure 5. Hence the number of acquisition of morphemes are significantly increased in both male and female Bengali speaking children as their age increased from 3 to 5 years.

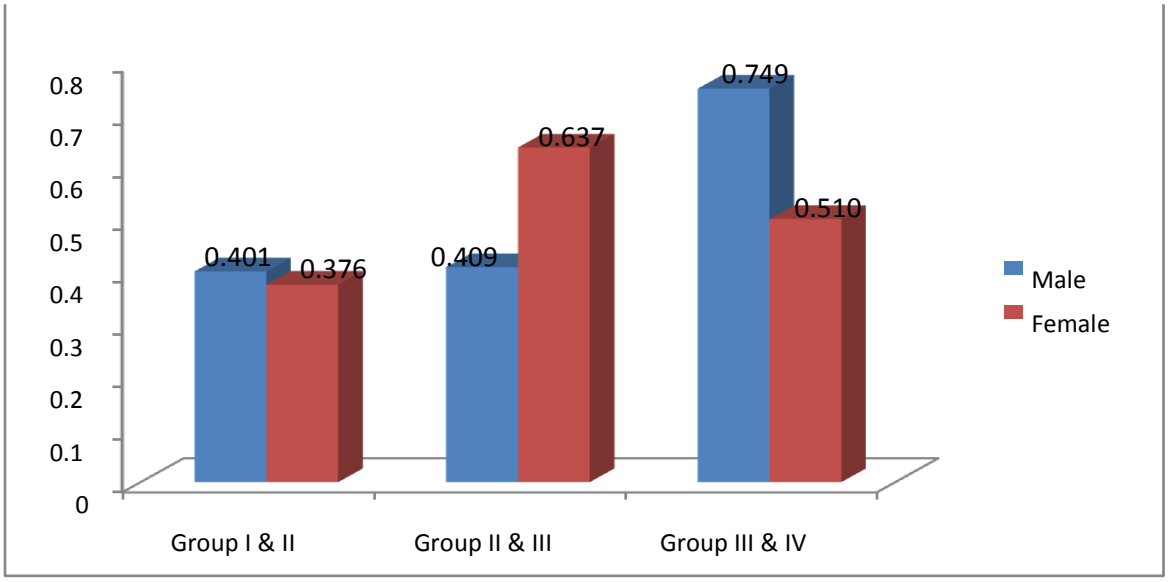


Figure 6: Comparison between group I & II, group II & III, group III & IV on the basis of MLUw acquisition in male and female children.

In the tables 4 in the result section, the mean difference is 0.401 between age group I and II which indicates that male children of 3 years to 3 year 6 months are significantly lower in acquiring words than the male children of 3 year 6 months to 4 years and similar findings was obtained in female children with the mean difference 0.376.

The mean difference between the age group II and III which indicates that the male children of 3 year 6 months to 4 years are significantly lower in acquiring words (mean difference is 0.409) than the male children of 4 years to 4 year 6 months and similar findings was obtained in female children with mean difference 0.637.

The mean difference between the age group III and IV also indicates that there is significantly lower number of words acquired in male children in the age range of 4 years to 4 year 6 months than 4 year 6 months to 5 years with mean difference 0.749 and similarly significant difference was obtained in female children with mean difference 0.510.

These findings are depicted in figure 6. It may be reported that the number of acquisition of words are significantly increased in both male and female Bengali speaking children as their age increased from 3 to 5

years.

Increase in acquisition of morphemes and words with increase in age was reported by Miller and Chapman (1981); Kazemi and Klee (2015)[11]; Eisenberg, Fersko and Lundgren (2001)[12]. Therefore, the age range 3 to 5 years is very crucial in the acquisition of morpheme and words in Bengali speaking children also.

9. SUMMARY AND CONCLUSION

The main purpose of the study was to find out the language acquisition among male and female children of 3 to 5 years. The duration of 2 years of age range has been divided into four groups. Each group has duration of 6 months. The participants included for the study belonged to the middle socio-economic background. The speech sample was collected through elicited speech in response to event specific and culture free materials like The Park, The Zoo, The railway station etc as well as through spontaneous speech and storytelling tasks. Each recorded sample was transcribed to MLUm as well as MLUw.

The findings of the current study provide better understanding of MLU in Bengali speaking children. The MLU scores obtained in the morpheme as well as in the words.

- The mean MLUm for Bengali speaking male children were obtained for the age group I, II, III and IV and they are; 2.81, 3.22, 3.8 and 4.1 respectively.
- The mean MLUm for Bengali speaking female children were obtained for the age group I, II, III and IV and they are; 3.4, 3.6, 4.2, and 4.9 respectively.
- The mean MLUw for Bengali speaking male children were obtained for the age group I, II, III and III and they are; 1.8, 2.2, 2.6 and 3.3 respectively.
- The mean MLUw for Bengali speaking female children were obtained for the age group I, II, III and IV and they are; 2.2, 2.6, 3.2 and 3.7 respectively.
- The acquisition of MLUm is more in length in Bengali speaking children than acquisition of MLUw.
- Gender differences were found statistically significant in the study which shows that the Bengali speaking female children surpass the acquisition of MLUm as well as the MLUw than the Bengali speaking male children across all the age groups.
- The development in the number of the MLUm and the MLUw in Bengali speaking children is increasing as age increases.

The MLUs obtained in morphemes and words for the age range 3 years to 5 years Bengali speaking children may be used for clinical purpose, clinical research and determining prognosis as it can serve the purpose of normative. The current study may provide a strong baseline for the future studies.

Longitudinal researches may be undertaken to assess the individual difference in development of Bengali morphemes and words which will substantiate the present study. This design of research can be used in higher age groups to know the language development and the plateaus reached in the acquisition of morphemes and words in Bengali speaking children. Similarly, language development may be explored

further in Bengali speaking children of lower age groups as from the review of literature it is evident that the children's goes through a period of transition from word to combination of words and acquisition of types of morphemes begins below 3 years of age.

REFERENCE

1. Brown, R. (1973). A First Language: The Early Stages. Cambridge, Mass.: Harvard University Press.
2. Miller, J. F., & Chapman, R. S. (1981). The Relation between Age and Mean Length of Utterance in Morphemes. *Journal of Speech Language and Hearing Research*, *24*(2), 154.
3. Dessai, R. D., & Karanth, P. (2009). Mean Length of Utterance and Syntax in Konkani. *Language in India*, 2.
4. Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*(3), 365-376.
5. Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language*, *18*(03), 553.
6. Snow, C. E., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J., & Vorster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research*, *5*(1), 1-20.
7. Thordardottir, E. T., & Weismer, S. E. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language*, *18*(52), 1-32.
8. Schachter, F. F., Shore. E., Hodapp, R., Chalfin, S., & Bundy, C. (1978). Do girls talk earlier? Mean length of utterance in toddlers. *Developmental Psychology*, *14*(4), 388-392.
9. Hickman, L. (2000). Sex differences in the language development rates of two-year olds. Portland state university.
10. Voniati, L. (2016). Mean Length of Utterance in Cypriot Greek-speaking Children. *Journal of Greek Linguistics*, *16*(1), 117-140.
11. Kazemi, Y., & Klee, T. (2015) Mean Length of Utterance (MLU) in typically-developing 2:6-5:6 year-old Persian-speaking children in Iran. *Journal of research in medical science*.
12. Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The Use of MLU for Identifying Language Impairment in Preschool Children. *American Journal of Speech-Language Pathology*, *10*(4), 323.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Unintentional Voice Modulations in Real World Conversations and its impact in Automatic Speaker Recognition

Soma Khan, Joyanta Basu, Rajib Roy, Madhab Pal and Milton Samirakshma Bepari
Centre For Development Of Advanced Computing (Kolkata), India

ARTICLE INFO

Article history:
Received 12/03/2020
Accepted 05/08/2020

Keywords:
Intonation,
Conversations,
Automatic Speaker recognition,
Segment Recognition Rate,

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU
Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Conversations are the most common form of human interactive communication. Rhythm in read out speech and conversational speech is bound to be different as the last one conveys straight away expressions of our thoughts. Purposefully capturing and analyzing the same has major importance in some of the cutting-edge technology research areas like Text–To-Speech synthesis, Emotion recognition, Automatic Speaker Recognition, Natural Language Understanding and Generation, Dialog modeling and many more (Pallotti, 2007) (Raux & Eskenazi, 2009). Here in this work, with our indigenously collected data resources and comprehensive analysis, we find out experimentally that what kind of rhythm is captured through the naturally occurring unintentional voice modulations in conversational speech, how they vary and to what extent these modulations can actually change the way real world Automatic Speaker Recognition system works. Experiments are designed with transcribed real world data collected in house as well as outside environment to closely analyze intonation patterns, speech overlaps and speaker turns in two speaker conversation cases. An indigenously developed method for conversational speech speaker recognition is described and a metric Segment Recognition Rate (%SRR) is introduced for performance measure. Impact of unintentional voice modulations on automatic speaker recognition performance is presented considering different mismatched speaking scenarios of real world.

1. Introduction

Two types of speaking modalities are found in our everyday verbal communication. In formal communication (read speech, announcements, public speaking etc.), speakers deliberately try to avoid the instantaneous voice modulations as far as possible, in order to inform, influence, impress or entertain listeners by their best efforts. Whereas informal verbal communication, specially human-human conversations are rather uncontrolled, they are straightaway expressions of our thoughts. For interactive and spontaneous nature, real world conversations carry a wealth of information at different levels which could be used to design machine learning algorithms for speech technology based automated application development.

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***
Corresponding Author: Soma Khan
Email: soma.khan@cdac.in

Conversational speech speaker recognition from recorded audio clip or live interactions is such an important application for automated meeting tracking (Vinyals & Friedland, 2008) (Biagetti et. al, 2016) and security related audio applications (Bonastre et. al, 2003). Conversation contains both intentional as well as unintentional voice modulations. Unintentional modulations are naturally occurring voice variations and have several representations (Tull & Rutledge,1996) (Matveev, 2013) (Hanilci et al., 2013) in situations like physiological changes in vocal apparatus (cold, sore-throat, aging), speaking amidst changed environment (shivering in cold, shouting in crisis, high vocal effort in noisy background, whispering in presence of undesired listeners etc.), varying state of mind (swing of mood, emotion, nervousness) and presence of secondary tasks (fast or slow speaking for work at hand, physical or cognitive stress). Except aging, most of these are momentarily and non-predictable, reflect different rhythmic expositions and cause broad intra-speaker voice variations even within a single conversation. These factors altogether pose some major challenges for conventional speaker recognition practices.

1.1 Organization of the article

Rest of the article is organized as follows. Section 2 clearly articulates the research objectives followed by Section 3, describing the research background along with its need and application area with respect to related prior works. Section 3 also explains on some important attributes of conversational speech which are experimentally studied in the following sections. Section 4 elaborates the efforts on data collection set up and data preparation processes for conversation experiments. Section 5 elaborately presents the findings and discussions on experimental analysis of intonation patterns, speech overlaps and speaker turns in the collected conversational corpus and relates the same to conversational speech speaker recognition complexity. Section 6 describes a novel conversational speech speaker recognition methodology followed by a series of speaker recognition experiments in section 7, designed in different mismatched scenarios of real world to validate the experiment findings. Finally conclusions are summarized in section 8.

1.2 Research Objectives

Unintentional modulations in human voice and related variations are largely captured through deviations in different attributes of real world conversational speech, but are rarely being studied earlier from automatic speaker recognition perspective. In this work, therefore our aim is as following:

- to capture the unintentional voice variations in two speaker conversations by simulating data collection setups within various real world speaking scenarios (face-to face and over telephone)
- to investigate the nature and extent of the captured variations in terms of deviations in conversational attributes with experimental and statistical analysis
- to figure out effects of these voice variations in terms of associated mismatch conditions on state-of-the-art speaker recognition performance.

2. Background

Accuracy of automatic speaker recognition is known to degrade severely under different mismatched conditions. Factors responsible for these mismatches usually include those of

speaker dependent (due to changes in speaker's voice) and communication channel or environment related (Pelecanos & Sridharan, 2001) (Reynolds, 2003). Variations in a person's voice can happen for intentional or unintentional reasons. As found in literature reviews, whether from conversation or from single speakers' speech, however, the baseline research activities of mismatched speaker recognition remain confined to the Intentional voice modulations. Intentional voice modulations (like mimic, disguised, whisper, converted or synthesized speech) pose critical security threats on recognition process and hence need urgent research efforts to develop relevant counter measures (Kajarekar et al., 2006). There are other situations when conversations are being scrutinized for security and surveillance purposes as part of the routine safety measures for inland security and maintaining law enforcement activities, where unintentional voice modulations need to be analyzed very carefully. But the fact is that, Unintentional voice modulations are not very well studied so far probably for reasons like scarcity of research ready transcribed data from conversations and requirements of multidisciplinary research knowledge. Though, these variations are very common and more importantly quite frequent in our day-to-day spoken communication.

Most of the prior works on conversational speech automatic speaker identification have targeted meeting room discussions; where environment is static, topic is predefined, mostly formal language is used and speech overlaps are minimal adhering to general office etiquettes. Simulation of meeting conversations have been done in some of the standard datasets like ICSI meeting speech (Janin et al., 2004), AMI meeting corpus (Carletta, 2006) etc. to carry out related research work. But real world human conversations are far different than this as mostly they do not follow any particular subject. Even though a particular topic is given, conversations often wander around talks on thoughts, memories and situations linked with the topic; and this is why they exhibit a wide range of speaking styles, nuances and linguistic strategies. Focused effort is needed to exploit this information and apply the same for building robust real world speaker recognition systems minimizing mismatched conditions.

2.1 Attributes of Conversational Speech

To analyze real world conversations closely we have selected following three most common and purposefully relevant attributes (among others) of conversational speech that largely represent the unintentional voice variations while conversing.

- Free intonation: This is caused by uncontrolled pitch variations. Unless having a conscious effort to discuss a particular subject, as in most cases, real world conversations mostly involve expressions (out of related or instantaneous thought process) reflecting perceivable differences in emotion or mood changes. These cause greater variations of pitch increasing intra-speaker variabilities. To study the extent of the same, intonation pattern analysis is necessary.
- Simultaneous speaking: In conventional social behaviour there are acceptable styles and rates of interruption, but conversations in real world often do not follow the same resulting into speech overlaps. Though speech overlaps appear to be impatience in turn-taking, but it indicates very resourceful, active and engaged talking between participants. Overlapped speech detection is a known research problem under speaker diarization (Anguera, 2012). But here, duration and degree of speech overlaps is our concern of interest (from speaker recognition).
- Obvious speaker turns: People who are close to each other (like friends, family and lovers) do not need to complete everything they initiate in conversations because their listeners are finishing it off in their own minds. Incomplete sentences are a sign of very fluent and engaged conversations but gives way to frequent speaker turns within short intervals. Though speaker turns are obvious in a conversation,

duration and frequency of such speaker turns are crucial for automatic speaker change detection and therefore important to consider in automatic speaker diarization and speaker recognition performance.

Other properties of real world conversations like usage of adverbials, continuity words and phrases, implicatures (context based implications), confirming questions and connected expressions, backtracking, monitoring talk etc. are also interesting, but we believe that the above three attributes are more affecting, frequent and thus relevant from conversation speech speaker recognition perspective.

3. Experimental Setup

Experiments are designed to collect relevant data and analyze the conversations. Following activities were taken up as experiment procedure:

Resource arrangement: Resources for data collection and experiments include in-house manpower, speakers to participate in data collection, necessary documents like appropriate text for read out speech (to compare with conversational speech), speaker and conversation metadata sheets, hardware for recording (desktop PC or laptop, headset, handheld microphones and mixer for face to face conversations, mobile phones for telephone conversations etc.) and software tools for speech data analysis and editing.

Data collection setup: To simulate all possible real world speaking scenarios and ease of comparative results analysis, experiments are designed for both readout and conversations considering environment conditions that of inside house as well as outside within real world. Inside house experiment setup was arranged at a relatively vacant corner of a 20ft X 16ft sized working laboratory within low to moderate level surrounding noise of fan, AC, distant babble, phone ringing etc. Recordings in speech studio are willfully avoided to discard evidences of practically unreal 'no noise' condition. Prospective speakers and speaker pairs for conversation are selected maintaining specific criteria of age (15 years to 60 years), spoken language knowledge, gender or gender pairs (like Male-Male or MM, Female-Female or FF, Female-Male or FM), area of mutual interest and work background. Once arrangements for all the necessary resources (manpower, speakers, hardware, software, documents) were done, speech data collection process was initiated in three different tracks, namely single speaker's readout, two speakers' face to face and telephone conversations. Data collection details are shown in table 1.

Speech data type	Read out	Face to face conversation	Telephone conversation
Environment	In house	In house, road side	Outside in real world
Device	Headset mice	Handheld microphone	Handset recorder
Sampling encoding	22050 Hz, 16 bit mono	22050 Hz, 16 bit mono	8000 Hz, 16 bit mono
Record Format	Microsoft wave PCM	Microsoft wave PCM	.amr, .mp3, .3gpp, .mp4
Duration	2 minutes	5 minutes	5 minutes
Speakers	36 (19M, 17F)	Same 36	Other 44 (20M, 24F)
Number	36 waves	40 conv	40 conv

Table 1: Specification of the collected data for experiments

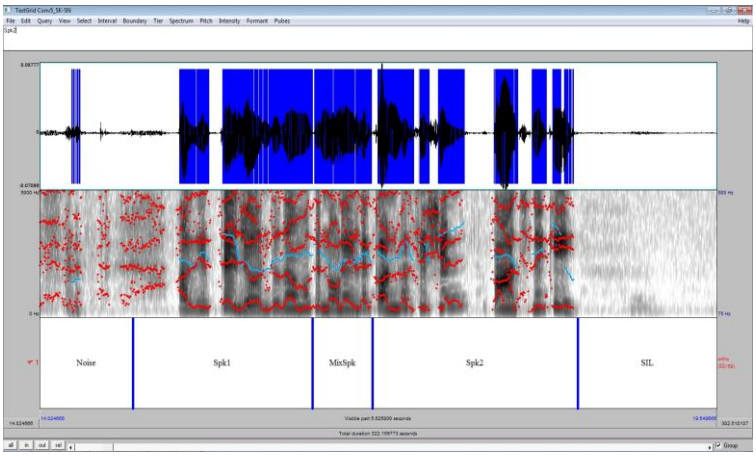


Figure 1: Screenshot of manual transcription of conversational speech data using Praat software

Data transcription: All the recorded conversations are transcribed using Praat software (open source speech editing software, available at www.praat.org). Figure 1 shows the process of manual transcription using Praat.After careful listening and seeing each conversation file, human transcribers have marked start-end boundaries for different events and labelled them as environmental noise (Noise), pure silence (SIL), overlapped speech (MixSpk) and speaker specific speech (Spk1 or Spk2 with vocal sounds)..

4. Experimental Methodology for Conversation analysis

4.1 Intonation pattern analysis

After verification of the transcriptions of face to face conversations, speaker specific pure speech (no overlap) portions have been elicited and merged (based on same label name) automatically to get individual speaker's speech profile which were used later for model creation. Pitch (F0 values) have been extracted from these profiles and readout speeches to get Pitch Occurrence Frequency Distribution (POFD) within a frequency span of 60 Hz to 420 Hz (adhering to pitch range of normal speakers aged between 15 years to 60 years) using bins of 20 Hz.

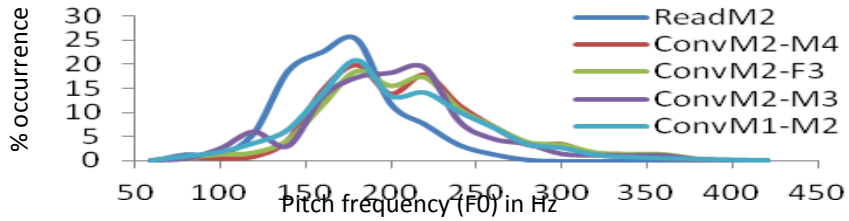


Figure 2: POFD of male speaker M2, Read vs Conversation, in-house

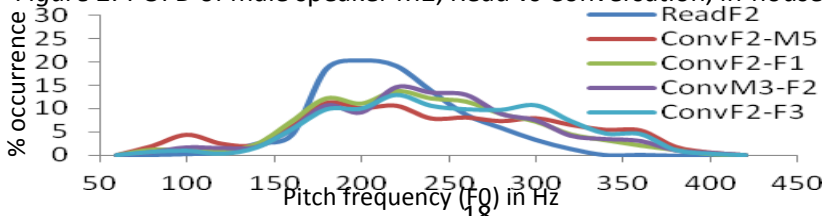


Figure 3: POFD of female speaker F2, Read vs Conversation, in-house

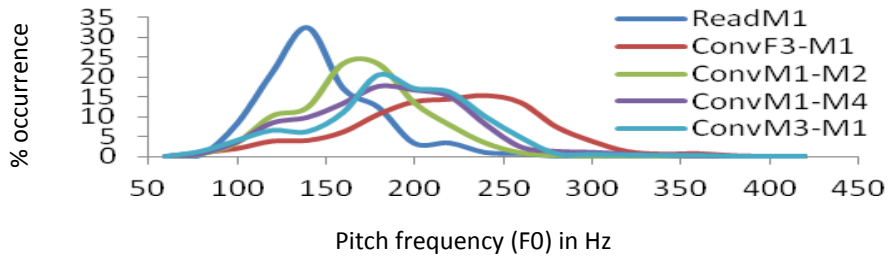


Figure 4: POFD of male spk M1, Read vs Conversation, some outside

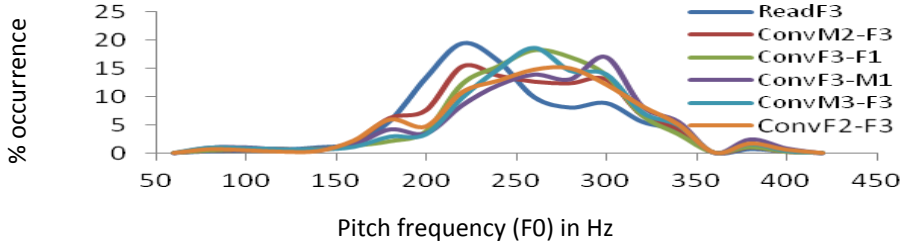


Figure 5: POFD of Female spk F3, Read vs Conversation, some outside

Figure 2 and figure 3 represent the POFD plots of read speech (indigo blue line) and face to face conversations of one male (M2) and one female speaker (F2) respectively, recorded all within in-house. Whereas figure 4 and figure 5 show similar POFD plots of another male speaker (M1) and female speaker (F3) respectively; but including that of some outside recorded (roadside garden area) conversations. Difference between readout speech and conversational speech intonation pattern, of same speaker is clearly observable in all the four figures.

In readout speech, pitch values have concentrated within certain pitch ranges, whereas greater variations are found in conversational speeches of both male and female speakers. More surprisingly, in-house conversations (shown in figure 2 and figure 3) have formed nearly similar POFD patterns though having differences in topic, speaker pair and gender pairs. But no specific POFD pattern is formed (in figure 4 and figure 5), when outside recorded conversations (like ConvF3-M1, ConvM1-M4, ConvF3-F1, ConvM3-F3) are considered along with. In summary, intonation pattern analysis suggests that read speech vs. conversational speech is bound to form a well distinguishable mismatched scenario in conversational speech speaker recognition because of huge intra-speaker variations formed by naturally occurring uncontrolled pitch modulations of same speaker in all real world conditions.

4.2 Speech overlap analysis

From all the studied conversations, overlap labeled speech segments were elicited to analyze the occurrence frequency percentage of overlap durations with respect to Male-Male (MM), Female-Female (FF) and Female-Male (FM or MF) conversations separately. These have been plotted in figure 6.

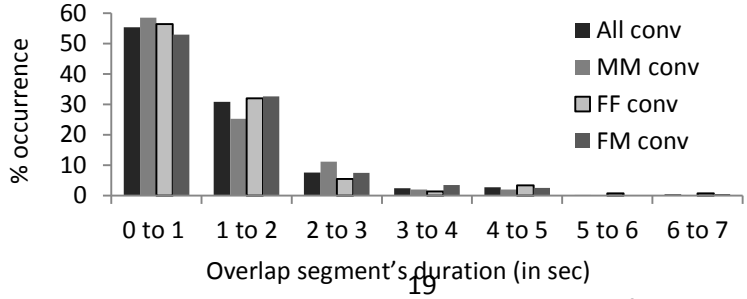


Figure 6: Speech overlaps duration with occurrence frequency (%)

From figure 6, it is visible that, across all the conversation types, overlapped speech segments up to 1 second duration are occurring in maximum numbers (about 52% to 58% of all occurrences), followed by those between 1 to 2 seconds duration (about 25% to 32%) and 2 to 3 seconds duration (about 5% to 11%) respectively. Occurrences of longer duration speech overlaps are almost negligible. Thus, small duration speech overlaps though indicate very co-operative talk, but pose challenge for automatic overlap detection.

Now, degree of speech overlap needs to be analyzed. From transcription of each conversational speech data, label specific segments of different speech events were accumulated to calculate duration proportion (in percentage) of pure speech (zero overlap), overlapped speech and nonspeech (silence and environmental noise) portions. Based on the presence of overlaps and nonspeech events, best case (overlap of minimum duration in ConvM3-F3 recorded at road side garden) and worst case (overlap of maximum duration in ConvF2-F3 recorded in house) speaking scenarios have been decided and presented in figure 7, 8, 9 and figure 10, 11, 12 respectively in terms of respective duration percentages.

Figure 7 is the actual speaking scenario of ConvM3-F3 as per manual transcription, where the portions (% duration) of pure speech of male speaker M3, female speaker F3 and overlapped speech is 53%, 24% and 23% respectively. After applying automatic speaker recognition, the resultant speaking scenario for the respective speech portions are 50.67%, 20% and 29% as presented in figure 8, where miss recognitions and true speaker in overlaps are included within 29%. After rechecking of the recognition results as shown in figure 9, it is seen that, out of this 29%, system has correctly identified male speaker M3 and female speaker F3 (from overlapped speech) in 14.67% and 4% occurrences leaving remaining 10.67% miss recognitions in ConvM3-F3.

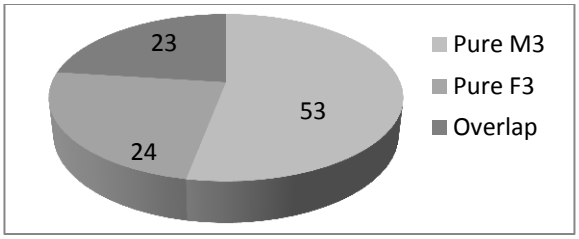


Figure 7: Actual speaking scenario, ConvM3-F3

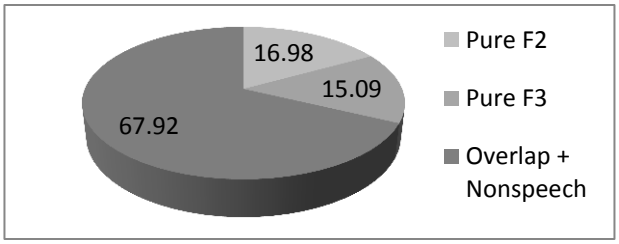


Figure 10: Actual speaking scenario, ConvF2-F3

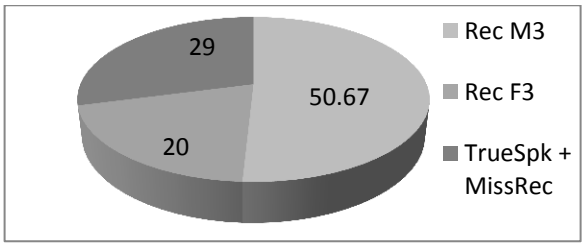


Figure 8: Speaking scenario as auto recognition

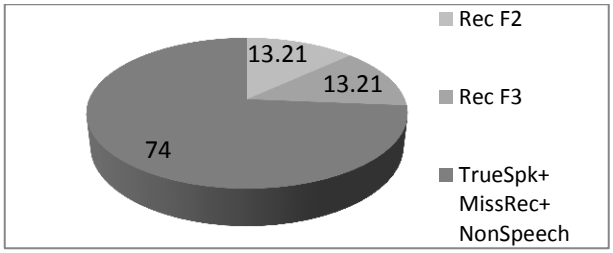


Figure 11: Speaking scenario as auto recognition

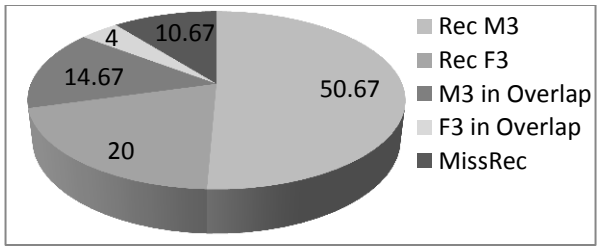


Figure 9: Speaking scenario after overlap analysis

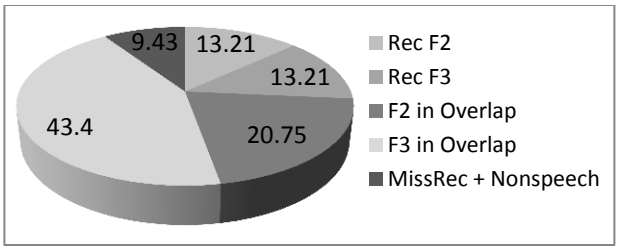


Figure 12: Speaking scenario after overlap analysis

In contrary, minimal number of pure speech events and more number of overlaps and nonspeech events are found in ConvF2-F3 making it a perfect replica of real world (though recorded within in-house) conversation or informal spoken communication. As shown in figure 10, respective portions (% duration) of pure speech of female speakers F2, F3 and that of overlap with nonspeech events are 16.98%, 15.09% and 67.92%, referring to actual manual transcriptions. After applying automatic speaker diarization and recognition, the resultant speaking scenario become 13.21%, 13.21% and 74% respectively as presented in figure 11. Within this 74%, segmental durations of nonspeech (mostly laughing, vocal sounds etc.), miss recognitions and true speakers in overlap cases were merged. Interesting observations were found after overlap analysis on the system provided recognition results. As depicted in figure 12, within the 74% non-pure speech portion, system has correctly recognized speaker F2 and F3 (as mostly spoken or dominant speaker within overlap) with 20.75% and 43.4% occurrences leaving only 9.43% actual miss recognitions or errors in ConvF2-F3.

Thus apparently looking worst case can even become one of the best cases out of real world conversation speaking scenarios. So, overlap analysis suggests that, degree of overlap (like partial, speech with laugh or babble, crosstalk, unintelligible etc.) actually determines the chances of miss recognitions and is obvious to matter in automatic speaker recognition, but it is not certain for duration of the same. Though, total duration of pure speech parts in conversations is important for model creation while system training.

4.3 Speaker turn analysis

For information sharing or carrying out the discussion, turn taking between speakers is obvious within every real world conversation. But duration and frequency of speaker turns can affect the performance of conversational speech speaker diarization and recognition in certain situations. In absence of facial expression cues like that of in telephony conversations, speaker turns could be more impatient or reluctant. To capture all such scenarios, label specific segmented speech data have been programmatically extracted from all face to face and telephony recorded conversations, within which, only speaker labelled pure speech segments are selected here to calculate the turn durations and related frequency for each conversation.

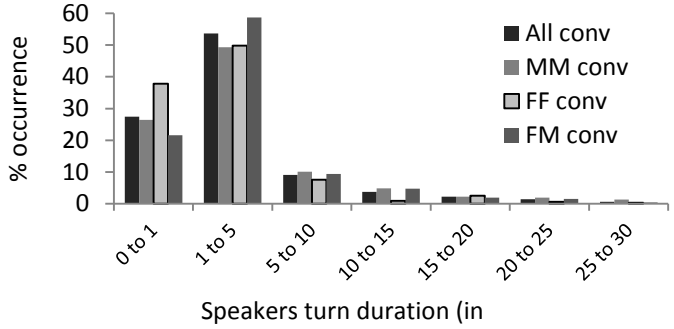


Figure 13: Speaker turn durations with occurrence frequency (%)

Speakers turn durations along with percentage frequency of occurrences have been plotted in figure 13. Turn durations of 1 to 5 seconds are found to be most frequent (about 53% to 58% of total occurrences) within all the studied conversations. That also hold true for all the MM, FF and FM conversations as well. More importantly, small duration speaker turns of even below 1 second are found to be next more frequent (about 21% to 37%) and they have occurred in

slightly more numbers specifically in FF conversations. In contrary, lengthier speaker turns of beyond 5 seconds duration are found with surprisingly less number of occurrences even below 10%.

From speaker turn analysis, thus it is being confirmed that speaker recognition from real world conversational speech have to perform mostly on short utterances of about 5 seconds in general and sometimes that might be even below 1 second duration. FF conversations are likely to be more challenging for automatic speaker change detection as well as recognition, as small duration and impatient speaker turns are found to be more frequent in real world FF conversations than that of MM or FM conversations.

5 Speaker Recognition Methodology

Here, both training and testing processes are performed on conversational speech automatically. Hence, an in-house developed Praat based automatic segmentation (first step of speaker diarization process) is introduced prior to automatic speaker recognition. This excludes purely nonspeech events (like field pauses & impulsive noises) and helps in easy elicitation of speech specific segments in conversations. Speaker specific acoustic characteristics are extracted from the speech segments in terms of short term spectral features, i.e. Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC), Log Area Ratio (LAR) and Partial Correlation (PARCOR) coefficients. In both the training and testing phases, pitch or F0 values are extracted from all the speech segments and POFD is calculated and saved as in section 5.

In training phase, speech segments having a single unimodal and similar POFD (same start & stop values with ± 20 Hz tolerance on both ends) are considered as the pure speech of same speaker and hence clustered together to form individual speakers' speech profile. Popular and stable unsupervised learning methods, both Vector Quantization VQ and Gaussian Mixture Models GMM are applied on the features extracted from individual speaker's speech profiles to minimize the effect of data redundancy and compact representation of speaker models. Speaker discriminative weighting is applied on the speaker models to assign larger weights on speaker discriminative features.

During testing, a novel Pitch Based Dynamic Pruning (PBDP) (Khan et al., 2012) technique is applied to prepare a list of most likely background (or candidate) speakers using POFD of train and test data. Finally, matching is performed for each test speech segment, with all the survived (after PBDP) speakers' models to get the highest matching scorer speaker as the identified speaker. After matching is done for all such segments of a test conversation, speaker identified in maximum number of segments is ranked as the first target speaker and the same with next highest number is ranked as the other target speaker.

6. Speaker Recognition Experiments

A number of experiments are conducted under matched and mismatched train-test conditions to see the impact of unintentional voice modulations on performance of conversational speech speaker recognition. As conversations have been processed in terms of segments, a new performance measure namely Segment Recognition Rate (SRR) in percent (%) has been introduced here.

$$\% \text{ SRR} = ((\#\text{SegTSpk1} + \#\text{SegTSpk2}) / \#\text{TotalSeg}) .100$$

Where, #SegTSpk1 and #SegTSpk2 represent numbers of correctly recognized speech segments of target speaker 1 and target speaker 2 respectively and #TotalSeg represent the total number of test segments generated (by automatic segmentation) from a trial conversation. Experiments are performed with MFCC, LPCC, LAR and PARCOR features, within different matched and mismatched speaking scenarios as described in sections 7.1., 7.2 and 7.3 respectively.

6.1 Read speech train, Conv. speech test, face to face

Here, system has been trained with nearly 2 minutes (as in table1, section 4.) of read speech data (per speaker) from recording of an English text passage of 15 sentences on a given topic. Results of testing on face to face conversations by the same speakers are shown in figure 14. For all type of conversations, best average SRR is found to be around 40% with LAR features which is still below acceptance level. FF conversations scored better result, but MM conversations are the worst hit of this mismatched speaking situation.

6.2 Conversational speech train, Conversational speech test, face to face

Speaker models were trained (as described in section 6.) from face to face conversations and unseen conversations of same speakers were given for testing. As shown in figure 15, this results into good average %SRR this time i.e. above 80% across all the conversation types. FF conversations yield better results.

6.3 Conversational speech train, Conversational speech test, over telephone

Similar experimental data is prepared here, but that is taken from real world telephony conversations recorded on handset recorder. Here, as per figure 16, average SRR is found to be around 60% across all the feature types, but this time, MM conversations scored good results with around 70% SRR.

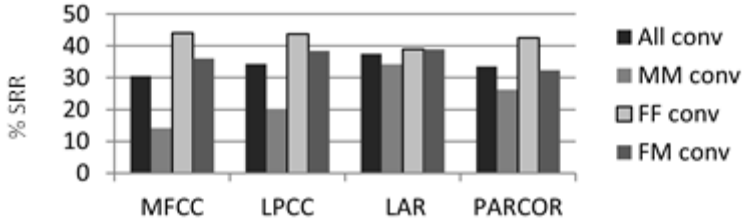


Figure 14: Read speech train vs. Conversation test, face to face

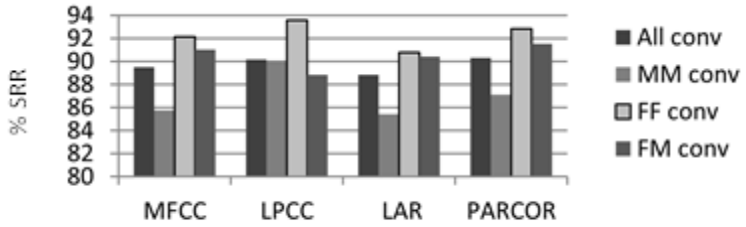


Figure 15: Conversation train vs. Conversation test, face to face

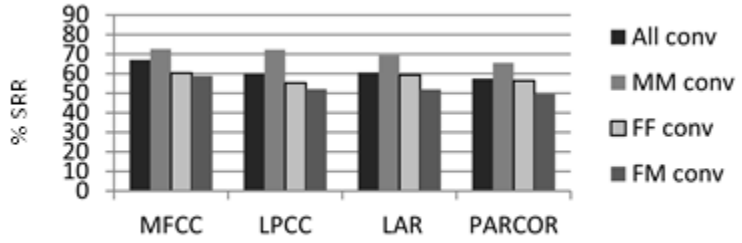


Figure 16: Conv. train vs. Conv. test, on telephone, real world

Thus, some important observations noted from the above experiments. Sample speaker's speech data taken from a conversation should be matched against speaker models trained with conversational speech only to yield purposefully relevant and comparable results. Prosodic variations present naturally in female speech contributed to form good speaker models which results into relatively less number of miss recognitions and good %SRR within less noisy in-house environment even under mismatched speaking scenario. Whereas low loudness, high speech rate, frequent speaker turns and mixing of non-speech and high frequency environment noises have affected female speech mostly in real world telephony conversations.

7 Conclusion

Automatic speaker recognition from conversational speech is quite different than that of from single speaker's fixed, predefined, prompted or co-operative speech. The first one needs to consider naturally occurring unintentional voice modulations, realized largely through deviations in various features of conversations. Effects of such deviations are clarified by experiments on two speaker conversations in terms of intonation pattern, speech overlap and speaker turn analysis. Both the analysis results and system recognition performance under different mismatched situations indicate to treat real world conversations by their types and recording background. Applying prior knowledge of gender pairs and environment may improve the recognition results. However, more data needs to be analyzed to make concrete decisions; though properly transcribed standard databases of real world two speaker conversations are not publicly available. This study is an effort to understand real world conversations more closely, to invent and apply robust methods for conversational speech speaker recognition handling associated voice modulations more gracefully.

References

Anguera, X., Bozonnet S., Evans, N., Fredouille, C., Friedland, G., & Vinyals, O. (2012). Speaker diarization: A review of recent research. *IEEE Transactions on Audio, Speech and Lanuage Processing, Feb. 2012, 20*(2). 356–370. https://ieeexplore.ieee.org/document/6135543

Biagetti, G. , Crippa, P. , Falaschetti, L. , Orcioni, S. & Turchetti, C. (2016). Robust Speaker Identification in a Meeting with Short Audio Segments. *Intelligent Decision Technologies, part of Smart Innovation, Systems and Technologies book series SIST, 57*, 465-477. https://doi.org/10.1007/978-3-319-39627-9_41

Bonastre, J. F., Bimbot, F., Boe, L.J., Campbell, J. P., Reynolds, D. A., & Magrin-Chagnolleau, I. (2003). Person authentication by voice: A need for caution. *8th European Conference on Speech*

Communication and Technology, Eurospeech 2003. (pp. 33-36). https://www.isca-speech.org/archive/eurospeech_2003/e03_0033.html

Carletta, J. (2006). Announcing the AMI Meeting Corpus. *The ELRA Newsletter January-March, 11*(1), 3-5. *AMI Meeting Corpus*. [Data set]. http://groups.inf.ed.ac.uk/ami/corpus/

Hanilci,C., Kinnunen, T., Saeidi, R., Pohjalainen, J., Alku, P., & Ertas, F. (2013). Speaker Identification from Shouted Speech: Analysis and Compensation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*. (pp. 8027-8031). https://doi.org/10.1109/icassp.2013.6639228

Kajarekar, S. S., Bratt, H., Shriberg, E., & Leon, R. D. (2006). A Study of Intentional Voice Modifications for Evading Automatic Speaker Recognition. *2006 IEEE Odyssey-The Speaker and Language Recognition Workshop*, (pp. 1-6). https://doi.org/10.1109/odyssey.2006.248123

Khan, S., Basu, J., Bepari, M. S., & Roy, R. (2012). Pitch based selection of optimal search space at Runtime: Speaker Recognition Perspective. *4th International Conference on Intelligent Human Computer Interaction (IHCI 2012)*. (pp. 292-297). https://doi.org/10.1109/ihci.2012.6481822

Matveev, Y. (2013). The Problem of Voice Template Aging in Speaker Recognition Systems, M. Zelezny et al. (Eds.), *International Conference on Speech and Computer – SPECOM 2013, Lecture Notes in Artificial Intelligence (LNAI)*, 8113, 345–353. Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-01931-4_46

Pallotti, G. (2007).Conversation Analysis: Methodology, machinery and application to specific settings. In H. Bowles & P. Seedhouse (Eds), *Conversation Analysis and Language for Specific Purpose* (pp. 37-52). Bern: Peter Lang. ISBN:978-3-0343-0045-2.

Pelecanos, J., & Sridharan,S. (2001). Feature warping for robust speaker verification. *In Proceedings of 2001: A Speaker Odyssey, The Speaker Recognition Workshop*, Greece. (pp. 213–218). https://pdfs.semanticscholar.org/f39d/0c70e0f89296045f0f11d79e15fa978e3081.pdf?_ga=2.209371404.103145289.1583406594-179049289.1559110441

Raux, A., & Eskenazi, M. (2009). A finite-state turn-taking model for spoken dialog systems. *NAACL '09 Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. (pp. 629-637). https://doi.org/10.3115/1620754.1620846

Reynolds, D.A. (2003). Channel robust speaker verification via feature mapping. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*. (pp. 53–56). https://doi.org/10.1109/icassp.2003.1202292

Tull, R.G. & Rutledge, J. C. (1996). Analysis of 'cold-affected' speech for inclusion in speaker recognition systems. *The Journal of the Acoustical Society of America*, *99*(4), 2549-2574. https://doi.org/10.1121/1.415166

Vinyals, O., & Friedland, G. (2008). Towards Semantic Analysis of Conversations: A System for the Live Identification of Speakers in Meetings. *IEEE International Conference on Semantic Computing*. (pp. 426-431). https://doi.org/10.1109/icsc.2008.58

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

TIMBRE BASED STYLE IDENTIFICATION

Kaushik Banerjee, Anirban Patranabis, Ranjan Sengupta and Dipak Ghosh
Jadavpur University, India

ARTICLE INFO

Article history:
Received 14/03/2020
Accepted 05/08/2020

Keywords:
Indian Classical Music,
Raga,
Timbre analysis,
brightness,
irregularity,
acoustic analysis

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

This work is about timbre analysis of the sound signals of vocal sound of eminent vocalists in Hindustani music and is aimed at the identification of timbral characteristics of these musical signals in a musical progression. Among timbre parameters, brightness, irregularity, odd and even harmonics, irregularity among partials of the signal and spectral centroid are studied. This approach leads to an understanding of the styles of the artists based on the timbral changes

1. INTRODUCTION

American Standards Association (ASA 1960) made a definition of timbre as "Timbre is that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar." Jensen (2001) also commented on the ASA definition as "This definition defines what timbre is not, not what timbre is. Timbre is generally assumed to be multidimensional". Basic definition of timbre is still a debatable issue and may cause more confusion than clarity. Here lies the challenge of timbre research. Although timbre is an important terminology in a most common musical vocabulary as well as it is a crucial component in the terminology of hearing science, it seems, that a consensus on a precise scientific definition is still due to be achieved. The focus of this work however is on the spectral and temporal character of timbre. It points out the importance of particular sequences of timbral features as being a crucial information carrier for the purpose of sound classification and identification of

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Kaushik Banerjee
Email:sitar.kaushik@yahoo.com

style of vocalists in Hindustani music. The temporal sequence of change is considered as the timbral structure. The extracted features are an important recognition cue when identifying sounds. For this study, twenty songs sung by five singers covering four ragas namely Bhairav, Todi, Darbari Kannada and Mian-Ki-Malhar were taken for analysis. Only the aalap (only the vocal part of a rendition without any influence of accompanying instruments like table etc.) portion of various ragas sung by various eminent singers in Indian classical music were considered Digitization of the signal was done @44100 samples/second (16 bit/sample). Partial to partial characteristics were studied from Long term average spectra (LTAS) of each signals and these constitute the database for the study. The main goal is to identify the style of a vocalist timbre aspects.

2. Objective

Human brain with little training in Indian classical music can identify style of a singer but an amateur cannot do so. In this work our main objective is to find computationally the timbre characteristics of the musical signals that help in identifying style of a vocalist.

3. Timbre Parameter

Spectral Centroid of a sound is a concept adapted from psychoacoustics and music cognition. It measures the average frequency, weighted by amplitude, of a spectrum. The standard formula for the (average) spectral centroid (Park, 2004) of a sound is $C=\Sigma C_j / j$ where C_i is the centroid for one spectral frame, and i is the number of frames for the sound. The (individual) centroid of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes, or: $C_j = \Sigma(f_i a_i) / \Sigma(a_i)$**.** Spectral centroid is the centroid of the spectrum.

Even-odds energy is a measure for the energy distribution on even and odd harmonics and is related to the subjective sensation of fullness of a sound, another important attributes that assist in specifying timbre and it depends upon the ratio of odd to even numbered partials.

Odd parameters are defined as $ODD = \Sigma(a_{2k-1}) / \Sigma(a_k)$. To avoid too much correlation between the odd parameter and the tristimulus 1 parameter, the odd parameter is calculated from the third partial. A even parameters are defined as $EVEN = \Sigma(a_{2k}) / \Sigma(a_k)$. Since tristimulus 1 + odd + even equals 1, it is necessary to keep only one of the two relations (Jensen K).

Tristimulus method of analysis proposed by Pollard & Jansson (1982), where the spectral energy is divided into three bands and the energy level for each band is determined i.e. the tristimulus is also a descriptor for the spectral energy distribution. It measures the energy in the fundamental-, the next three partials, and the higher partials in relation to the whole energy. Since the sum of Tristimulus one, -two and -three equals "1" only two values need to be calculated. The tristimulus values have been introduced in (Pollard et al. 1982) as a timbre equivalent to the color attributes in the vision. They used it for analyzing the transient behavior of musical sounds and for classification of musical timbre. Tristimulus is defined by the following three equations: TRISTIMULUS 1 (T1) = $a_1 / \Sigma(a_k)$, TRISTIMULUS 2 (T2) = $(a_1 + a_2 + a_3)/ \Sigma(a_k)$ and TRISTIMULUS 3 (T3) = $(a_4+a_5+a_6……..+a_k) / \Sigma(a_k)$. where a_1 stands for the amplitude of fundamental or 1^{st} harmonic, a_2, a_3, a_4 are the amplitudes of 2^{nd}, 3^{rd} and 4^{th} harmonics respectively and k is the partial index. So T1 is the descriptor for the spectral energy distribution of fundamental, T2 is the descriptor for the spectral energy distribution of next three harmonics and T3 is the descriptor for the spectral energy distribution of higher partials.

Irregularity/spectral smoothness is defined as the sum of the amplitude minus the mean of the preceding, same and next amplitude in dB i.e. the local mean, are compared with the current

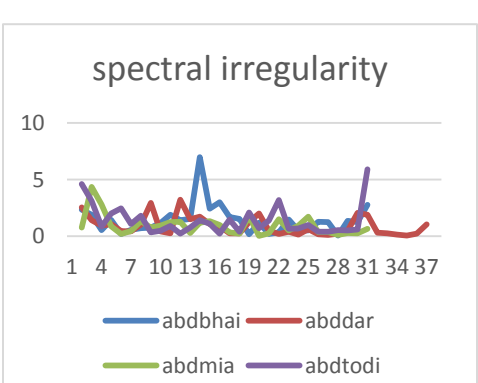
amplitude value. Bregman (Bregman 1990) remarks that the smoothness of a spectrum is an indicator for partials belonging to a same sound source and a single higher intensity partial is more likely to be perceived as an independent sound. It has also been found to be useful in revealing complex resonant structures of string instruments. It is written as IRREGULARITY = $\Sigma Ia_k-(a_{k-1}+a_k+a_{k+1})I/3$**,** Where a_k is the amplitude of k^{th} harmonic, a_{k-1} and a_{k+1} are the amplitude of previous and next of k^{th} harmonic. In this paper, an alternative calculation of the irregularity is used, where the irregularity is the sum of the square of the difference in amplitude between adjacent partials, IRREGULARITY = $\Sigma (a_k-a_{k+1})^2 / \Sigma a_k^2$ and the N+1 partial is supposed to be zero.

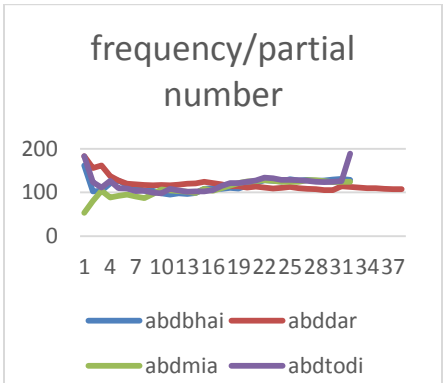
Brightness is one of the most important perceptive attributes that assist in specifying timbre. It is determined by the location of the mean of the energy distribution on the frequency continuum (spectral centroid). Brightness is calculated as the spectral centroid which is correlated with the subjective quality brightness. (Jensen K, 2002) The brightness is calculated as BRIGHTNESS = $\Sigma (k\ a_k) / \Sigma a_k$, where k is the partial index, and a_k is the amplitude of the k^{th} partial. A closely related attribute is sharpness. If the partial index multiplication k is replaced with the frequency of the partial, the brightness is expressed in Hertz. For harmonic sounds, this is equivalent to multiplying the brightness with the fundamental.

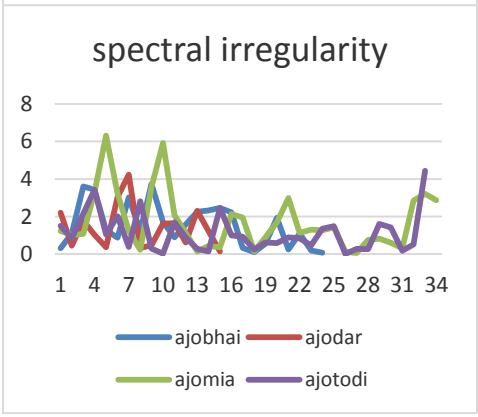
4. Results and discussion

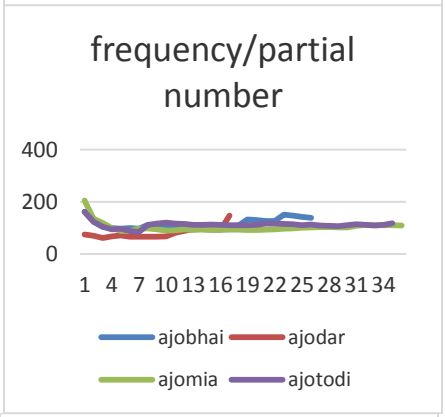
Table 1 shows the deviation of energy of each partial from its mean value (standard deviation of spectral irregularity) and average frequency per partial number. Singer ABD shows the similar behavior of spectral irregularity in raga bhairav and todi and also in raga darbari and mia ki malhar but the frequency per partial number is different for all the ragas. Similar behavior of spectral characteristics are observed in ragas bhairav, darbari and mia ki malhar is observed in singer AJO and hence that may be a style stamp of him. Spectral irregularity is almost similar in singer AMI, ANI and ARN's signals but frequency variations per partial number are wide apart. From figure 1 it is observed that the variations of spectral energy with partial numbers is less in the spectrum of singer ANI while it is more for singers AMI and ARN. It is also observed that the variations of frequency partial numbers is high in the spectrum of singer AMI and AJO while it is quite less for singers ANI and ARN. Mid frequency sounds have uniformity in timbre characteristics but the randomness in timbre increases with the increasing pitch for the signals of AJO and ABD. From table 2 it is found that the centroid of amplitude is equally spaced in ragas bhairav, mia ki malhar and todi of singer ABD, mia ki malhar and todi of singer AJO, bhairav, darbari and todi of singer AMI, bhairav, darbari and todi of ANI and bhairav and mia ki malhar of songer ARN. For all the singers T1 and T2 are less while T3 is high. This proves that the energy pumped up at the high frequency partials for all the signals. In most of the the vocal signals of AJO, ANI, ARN and AMI the uniqueness of sound lies in the multiple decay of the total energy as well as the periodic fluctuations of the harmonics. The waxing and waning stems from multiple decay of the higher harmonics is a speciality of these vocal sounds. This property is precisely responsible for constantly pumping energy to the higher harmonics, leading to a resonant frequency that is very different from other vocal signals. Low frequency partials also have higher energy for singers ABD and AJO. Amplitude fluctuation is a major criterion of their sound signals. It has also been observed that the sound spectrum of these vocal sound has very irregularly occurring bands of spectral peaks which sustain for long periods. This is due to lack of supply energy to the higher harmonics. It is also found that the centroid of frequency is equally spaced in ragas bhairav, and mia ki malhar of singer ABD, darbari and mia ki malhar of

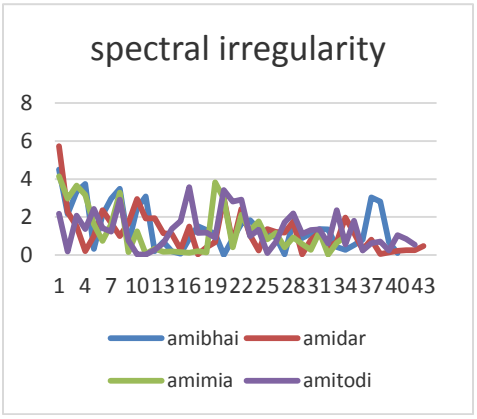
singer AMI, bhairav, darbari and todi of ANI and bhairav and todi of songer ARN. For all the singers odd and even partials have similar energy.

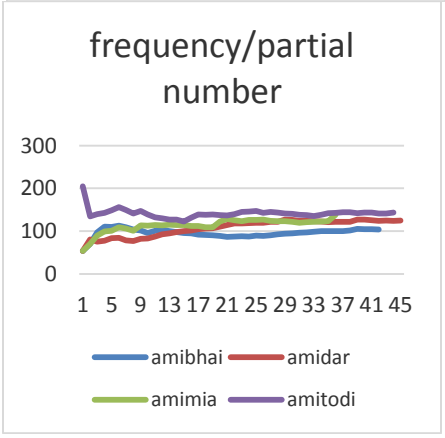












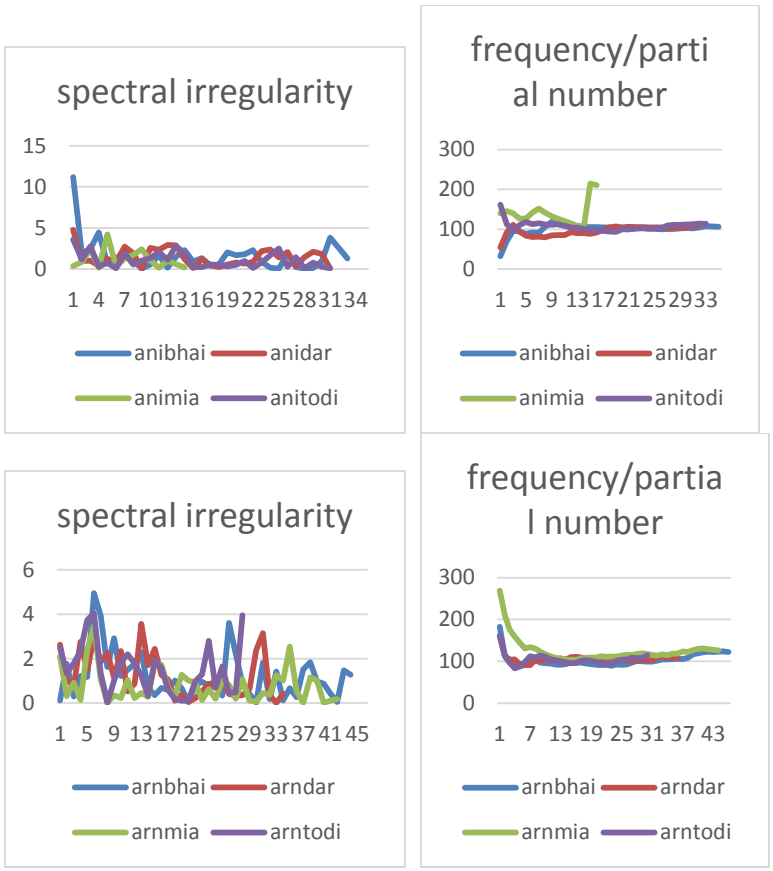


Figure 1: Spectral irregularity (partial number along X axis and spectral irregularity along Y axis) and frequency per partial number (partial number along X axis and frequency per partial number along Y axis) of five singers.

	Spectral Irregularity				Frequency per partial number			
	bhai	dar	mia	todi	bhai	dar	Mia	todi
ABD	1.305	0.847	0.896	1.335	14.481	16.037	17.729	20.316
AJO	1.152	1.1624	1.239	1.031	19.555	20.026	19.929	11.856
AMI	1.169	1.069	1.25	0.935	10.441	19.367	16.035	11.557
ANI	1.023	1.079	1.16	0.939	14.249	12.648	31.189	11.93
ARN	1.062	1.016	0.777	1.144	16.191	10.738	29.41	13.177

Table 1: Standard deviation of Spectral irregularity and frequency per partial number

	brightnes	T1	T2	T3	odd	even	irre 2	Centroid
abdbhai	16.74	0.04	0.1	0.87	0.46	0.5	0.05	62.56
abddar	19.72	0.03	0.07	0.9	0.47	0.5	0.04	58.32
abdmia	16.73	0.04	0.09	0.87	0.46	0.5	0.04	61.85
abdtodi	16.68	0.04	0.1	0.87	0.46	0.5	0.08	65.85
ajobhai	14.53	0.03	0.09	0.88	0.46	0.51	0.06	68.28

ajodar	9.76	0.05	0.14	0.81	0.49	0.46	0.09	52.34
ajomia	19.59	0.02	0.07	0.91	0.48	0.5	0.04	55.1
ajotodi	19.19	0.03	0.07	0.9	0.49	0.48	0.05	61.21
amibhai	23.34	0.02	0.05	0.93	0.48	0.5	0.06	54.17
amidar	25.88	0.02	0.04	0.93	0.49	0.49	0.05	68.58
amimia	20.29	0.03	0.06	0.91	0.47	0.5	0.07	68.42
amitodi	25.37	0.01	0.04	0.95	0.48	0.51	0.06	81.26
anibhai	19.38	0.04	0.05	0.91	0.48	0.47	0.08	57.26
anidar	18.56	0.04	0.06	0.9	0.49	0.47	0.07	57.65
animia	9.18	0.04	0.13	0.83	0.45	0.51	0.09	83.8
anitodi	18.18	0.02	0.07	0.91	0.49	0.49	0.05	58.21
arnbhai	24.72	0.02	0.07	0.9	0.47	0.5	0.04	57.66
arndar	20.4	0.02	0.06	0.93	0.47	0.51	0.06	60
arnmia	24.26	0.03	0.06	0.91	0.48	0.5	0.04	66.28
arntodi	16.8	0.02	0.08	0.9	0.47	0.51	0.06	57.72

Table 2: Timbre parameters

	brightnes	T1	T2	T3	odd	even	irre 2	Centroid
ABD	3	2	3	3	3	3	3	0
AJO	0	0	0	2	3	2	0	0
AMI	2	2	4	2	0	2	0	3
ANI	0	3	0	0	2	0	3	2
ARN	0	0	2	2	0	2	2	0

Table 3: K nearest neighbor of among four signals to each artists.

5. Conclusion

Timbre parameters were measured henceforth. Timbral characteristics such as tristimulus (T1, T2 and T3) and the odd and even parameters have been chosen in view of the energy distribution in partials, whereas spectral brightness and irregularity are descriptive of the harmonic content. Uniqueness of sound lies in the multiple decay of the total energy as well as the periodic fluctuations of the harmonics in all the signals except ABD. The waxing and waning stems from multiple decay of the higher harmonics is a speciality of all other vocal sounds. This property is precisely responsible for constantly pumping energy to the higher harmonics, leading to a resonant frequency that is very different from other vocal signals. Amplitude fluctuation is a major criterion of the sound signals. It has also been observed that the sound spectrum of some vocal sound has very irregularly occurring bands of spectral peaks which sustain for long periods. This is due to lack of supply energy to the higher harmonics. Timbre variation is observed for most of the vocal sound signals due to their variations in timbre properties. We also found some sound is highly inharmonic and its timbre property varies with the pitch. Sound of mid frequency sounds have uniformity in timbre characteristics but the randomness in timbre increases with the increasing pitch for AJO and ABD. Table 3 is self-explanatory. Style of artist ABD can be identified unanimously based on the Timbre characteristics.

References

Bregman, A. (1990). "*Auditory Scene Analysis: The Perceptual Organization of Sound*". MIT Press: Cambridge.

Jensen K and Georgios M (2001). "*Hybrid Perception*". 1st Seminar on Auditory Models, Lyngby, Datalogisk Institute, University of Copenhagen, Universitetsparken1, Copenhagen, Denmark.

Jensen, K. (2002). "*Perceptual and physical aspects of musical sounds*". University of Copenhagen, Universitetsparken1, Copenhagen, Denmark.

Park Tae Hong (2004). "*Towards Automatic Musical Instrument Timbre Recognition*", (PhD thesis) Department of Music, Princeton University.

Pollard H. F., Jansson E. V., (1982). "A tristimulus method for the specification of musical timbre". *Acustica*, 51, 162{171)

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

The cognitive aspect of word meaning in the *Brahmakāṇḍa* of *Vākyapadīyaṃ*

Rusa Bhowmik
Jadavpur University, India

ARTICLE INFO

Article history:
Received 22/03/2020
Accepted 06/08/2020

Keywords:
Pāṇini's Aṣṭādhyāyi, post-Pāṇinian grammar, linguistic philosophy, cognitive significance

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU
Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Vākyapadīyaṃ is a grammatical narrative written by *Bhatṛhari* around 5th CE. The concept of word in Indian grammatical tradition can be traced back to *Pāṇini's Aṣṭādhyāyi* which represents an analysis of the structure of Sanskrit grammar. *Vākyapadīyam* yields as one such treatise of post-*Pāṇini*an grammar. The research objective of this paper is to understand one aspect of the Indian grammatical tradition; i.e. word meaning from the lens of cognitive linguistics. The mind is perceived as the embodiment of language which is realized through its' internal identity and the consequence of individuality here is distinguished as a conventional norm for logical and linguistic philosophy. The application of the meaning based on the object will determine and presuppose the functionality of the type in the cognitive level. It is perceived by various schools of philosophy and grammatical traditions in various ways and aspects and hence it is devoid of a non-singular interpretation.

1. Introduction

Vākyapadīyaṃ is a grammatical narrative written by *Bhatṛhari* around 5th CE. It consists of three parts or as we can say *Kāṇḍas*. The three *Kāṇḍas* consist of *Brahmakāṇḍa*, *Vākyakāṇḍa*, and *Padakāṇḍa*. *Vākyapadīyaṃ* is also sometimes referred to as *Trikāṇḍī*. There are a few alternative names of two of the *Kāṇḍas*. The *Brahmakāṇḍa* is also known as *Āgamakāṇḍa* and *Āgamasamuccaya*. The *Padakāṇḍa* is known as the *Prakīrṅakāṇḍa* and *Prakīrṇakakāṇḍa*.

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***
This work is an outcome of a JU-RUSA 2.0 sponsored project entitled 'Vakyapadiyam and Pramanasamuccaya: A comparative study of the explanatory tools and techniques in the light of their respective socio-cultural perspectives' sanctioned to Dr. Samir Karmakar of School of Languages and Linguistics, Jadavpur University
Corresponding Author: Rusa Bhowmik
Email: bhowmik.rusa@gmail.com

Bhatṛhari composed a commentary on the *Brahmakāṇḍa* and *Vākyakāṇḍa* and it is known as *Vṛtti* (Sharma, pp. 1-2). *Bhatṛhari* in *Vākyapadīyaṃ* discussed his notions based on *Patanjali*'s *Vyākaraṇamahābhāṣya* which is again based on *Pāṇini's sūtras* and *Kātyāyana's vārttikas* (Sharma, p. 13).

The naming of *Vākyapadīyaṃ* is generally construed as the designation of all the three *Kāṇḍas* together. The first two volumes were meant to be the actual treatise and the final volume was mostly the commentary, and supplement to the first two volumes. Even records from I-tsing's testify that the third volume was considered as an independent book in their contemporary time. Later scholars argued that the second volume and the third volume is called *Vākyakāṇḍa* and *Padakāṇḍa* respectively, thus, *Vākyapadīyaṃ* derives its' name from all the three books; the first two volumes focus on the sentence and its' meaningful constituents, and the third volume deals with the overall commentary of the first two volumes. Eventually, it was later analysed that the three volumes of *Vākyapadīyaṃ* is named individually from the most important and significant constituent of the first sentence in each *Kāṇḍa*.

2. Research Objectives

The concept of word in Indian grammatical tradition can be traced back to *Pāṇini's Aṣṭādhyāyi* which represents an analysis of the structure of Sanskrit grammar. As mentioned earlier, *Vākyapadīyam* yields as one such treatise of post-*Pāṇini*an grammar. The research objective of this paper is to understand one aspect of the Indian grammatical tradition; i.e. word meaning from the lens of cognitive linguistics.

3. Theoretical Background

Grammar (*vyākāraṇa*) was considered as one of the six parts of the *Vedaṇgas*. The importance of grammar was due to its' indispensability in the understanding of the Vedic scriptures. *Śabda* has been denoted by scholars from the early Vedic period as a valid source of knowledge. *Śruti* and *smṛti* were two components of early Vedic teachings and this can only be derived from the verbal testimony of knowledge.

It has been debated in grammar whether the concept of verbal testimony (*śabdapramāṇa*) should be considered as immutable or mutable. Both the notions have been reflected and assimilated in the works of *Pāṇini* and the later grammarians such as *Bhatṛhari*.

The consciousness is based on the philosophical reflection on experience and the component is overtly expressed through the manifestation of perception. The epistemic analysis of meaning gulf between the objective and the subjective, uniting its' understanding with realization. The metaphysicality of the system brings forth the discussion of the concept of perception and inference. This was also one of the first pursuits towards the domain of linguistics.

The focus in the Indian epistemology and logic has always been on the dichotomy of the presence of the eternal and the transient. The significance of experience is constricted due to the surrounding forces and the substantiation towards the magnitude of cognitive self is reflected in the hierarchy between the two forms which is then transformed into the sacred and secular in a socio-historical context and is reflected in the texts, art, culture, and politics of the 5th CE.

The grammatical works were impelled with spirituality and *Vākyapadīyaṃ* is no different in this sense. The knowledge of self exists in the cognition which is reflected in the creation of the art in the consciousness. The mind is perceived as the embodiment of language which is realized through its' internal identity and the consequence of individuality here is distinguished as a conventional norm for logical and linguistic philosophy.

4. Discussions

The meaning through the understanding of language is expressed through the structure and its' constituents. The standing meaning is fixed by convention and the interpreter determines the context dependent meaning.

Śabda accepts and accommodates the concept of perception through an epistemic function of experience. The ontology and the epistemology of the function of *Śabda* is thoroughly discussed in *Vākyapadīyaṃ*.

(1) *anādinidhanaṃ brahma śabdatattvaṃ yadakśaraṃ*

vivartate'rthabhavena prakŗyā jagato yataḥ VP 1.1

The nature of *Brahma* is immutable which is identical with *Akśaraṃ* (the word), which is constituted as *Śabda Brahma*. *Bhatṛhari* had analysed grammar based on the philosophy and the doctrines of previous philosophers. Thus, the objective view refers to the evidentiality of the *Brahma* which is essentially consistent in nature. Verbal testimony (*śabdapramāṇa*) is the result of perception (*pratyakśaḥ*) and inference (*anumāṇa*). *Prakŗya* is a catalyst in the attainment of the realization of the *Śabda Brahma* and the production of the perception of the reality is in the reflection of the *vŗtti* (mental image). The cluster of components is formed to two variables; the external reality and the internal thought process.

Meaning is objective and its' form is expressed in its' realization. The experience of cognition through the realization of art is dynamic in nature and that helps in seeking the true meaning of knowledge. The proposition in relation to the one-dimensional or linear sense needs to be validated through philosophical investigation. The experience is described in lieu of the regular understanding of the world which is context sensitive or situation based entity and requires a presuppositional understanding to fit in the required knowledge. The stratum of mind is not always dissociated from the act of cognition rather combined of elemental matter and form.

The dynamic between the eternal and non-eternal reflects in consciousness. In the Indian tradition the bridge between the knowledge and the experience of the knowledge is essentially a dynamic process in nature. The embodiment of form is subjected to the logic of symbolism. The symbolism is an output of an idea which is limitless in nature and the expression of the idea is limited.

The concept of word meaning is associated with the philosophy of permanence and impermanence. It must be noted that the cognition of the object which is signified by a word is reposed on the speakers' mind than the object itself. The property of meaning becomes subjective with experience. The meaning of a word can be denoted by its universality along with its contextuality. The contextuality can be sub grouped again based on its inherent matter, i.e. a constant; and its interpretation. A word has its' own meaning and it is constant until it is referred from the perspective of the sentence. When it is said 'red'; the color 'red' has a standing meaning, i.e. a property which has been conventionally fixed. Depending on the context and usage, the meaning of the phrase 'red blood' and the 'red sky' differs.

The construct is real based on the perception of the experience of the mental state. The presence of meaning is embedded in words, but it is realized through sounds. The justification of meaning through the articulation of a string of linguistic symbols and that utterance in a particular sequence is preserved through the discourse. The representation of the concepts through speech is language. The mind does not represent ideas as they are, the universal concepts which are acknowledged are innate in nature and that existence is perceived by experience. The gap between the proximal and distal objects in the idea of the experience of perception observed in the conjunction between the existence and experience, therefore, the object realized and perceived is dependent on the cognitive ability.

The utterance of a word is dependent on the observation of the subject and the view of the object. The uniformity lies in the reception between the connection between words and experience. The external world is perceived as a reality that lives in the imagination and is transcended through the existence of the self. The totality of language is subsumed by its' linguistic units. The referential meaning of language is reflected in the meaning- expressing unit.

(2) *bāgrūpatā cedutkrāmedavabodhasya śaśvatī*

na prakāśaḥ prakāśeta sā hi pratyavamarśinī VP 1.124

The conceptuality of meaning is reflected in the principles followed by the spirit of a language, rather than any language. The attainment of the potency to understand the phase of language is not directly accessible. The functionality is abstracted by the two aspects; the utterance (*uccārita*) and the realisation of the utterance, i.e. the linguistic form (*śabdasvarupa*). The impression of a word is already established in the mind which is later realized through the linguistic unit, i.e. language through sound. The sound sequences form the meaning that establishes the connection to the particular feature of an entity.

Nonetheless, the external reality can be viewed to correspond to word meanings in reality. The understanding of the meaning is intended in the speech. The purpose of the speech is to convey the knowledge which is recognised in the cognition of the object than the object itself. Thus, it is compelled that the meaning of the word is embedded in the internal structure which is reflected by the speaker's intention.

The speaker understands the inherent along with the contextualized meaning of the word with reference to the sentence. The mental concept of the word is already present in the mind. Therefore, a meaning is not singular in nature, it is dependent on various factors and most importantly it is based on its' structure and the constituents. The concept of meaning is not real on the outside, rather its' existence lies in the mind. The intention in the speech of the speaker reflects the conception of meaning; the utterance reflects the reality of the speech.

The property or the attribute of the language is cognitively developed in a sequence of the understanding of the world. The communication is conceived in terms of the intended meaning of an utterance. The function of an utterance lies in its' semantic autonomy. Thus, the explication of the demonstration of the possibility of the outcome is dependent on the experience of the mind perceives and not just the reality that exists in the mind of the self.

The emphasis on the notion of consciousness conveys the intention of the thought through which it was processed. The nature of interrelation between mind and language is reflected through words. The correlation of the structure and form is based on the essence of the reality which is articulated through the act of communication.

(3) *ekameva yadāmnātaṃ bhinnaṃ śaktivyapākśyat*

apṛktve'pi śaktibhyaḥ pṛthaktveneva vartate VP 1.2

The presence or the absence of anything approves or denies a certain fact. But, this also denotes to accept or disprove any statement; one needs to have cognition of the concept. Whether it is accepted or disproved due to its' qualities or references comes into focus into a much latter part. This expresses the possibility of the existence or non-existence of a unit which needs an instrumental functioning of concept to argue or deny the expectancy.

(4) All Bengalis do not like fish.

(5) All Gujaratis do not like fish.

The concept of fish and the concept of Bengalis are primarily needed. Then, one need to understand the cultural significance of fish related to the Bengali community. Now, to deny that cultural correlation one needs to know the concept of the stereotype that is adhered by the general mass. Therefore, the meaning of the sentence needs to be separately understood by words, as Gujaratis do not like fish would make a paradigm shift and this won't presuppose the cultural context; again as a whole it would be a valid statement, but the thought process that one

arrives at denying a stereotype of Bengalis not liking fish is perceived as uncommon, even though statistically the current scenario may be different. The sense of the concept in the cognitive ability thus denotes words like Bengalis, and fish would separately mean a concept that can be independent of cultural context and stereotype as well as can be a heavily loaded sentence of cultural significance. The sentence meaning is represented as an indivisible whole. Word is the essence of that reality which is interpreted and realised in the verbal essence of the ultimate reality which is distinct in nature.

If we bring a word which has a meaning and attribute based on reality Y, then the meaning or rather the quality of x is embedded in Y. Y is a representation; an object, a token. The token is a representation in the reality, on the etic level. The meaning of Y is relative in nature. This is because meaning can be particular or universal. The theory of meaning lies either in its' descriptive form or explanatory form. The specificity of the token is dependent on the context. For example, a bold x, an italic x or an underlined x are all x'es. But are all x'es same or different? If the meaning of Y is dependent on x then all the x'es will be treated individually with respect to Y. But, if the meaning of Y is independent and will not be affected by the quality of x, then x in every form will be seen as the different facets of a singular entity x.

(6) Bold x = x1, italic x = x2, underlined x = x3

x= x1/x2/x3 or x is not equal to x1/x2/x3

x has 3 divisions x1,x2,x3

Therefore, Y has attribute x may mean Y has x1/Y has x2/Y has x3.

Now, the quality of x will determine Y's meaning and will lead to Y1 dependent on x1, Y2 dependent on x2 and so on. The application of the word in the context will determine the situation how a certain word's meaning will be constituted.

(7) Give the ANIMAL some FOOD to eat.

It is common knowledge that all animals eat food. But all animals does not eat same food. If one does not have a previous knowledge of which animal eats which food, the meaning construed will seem either funny or stupid or both.

(8) ANIMAL = {cow, lion, donkey, man}

(9) FOOD = {milk, meat, grass, water}

Though syntactically correct statement like 'give the lion some grass to eat' is possible along with 'give the donkey some meat to eat', but these sentences are absurd.

The meaning of the word and the reference along with it is equally important to understand the situation. A word's establishment is dependent on various factors, along with the word meaning (*padārtha*).

5. Conclusion

The totality of the phenomenal reality is postulated through individual words and elements expressed by the conceptual symbol. The reality in the implicit way of rational thinking is based on the reconstructed form of reality which is achieved through sounds. This can be achieved not by mere output of sounds through a mental concept, but this is deeply embedded in the grasp of its' entirety in the intellect and the theoretical aspect of understanding of the concept. The abstract form is always context-insensitive. The application of the meaning based on the object will determine and presuppose the functionality of the type in the cognitive level.

Thus, the essence of the reality is the being which is manifested and the manifestation of the illusions portrayed through the numerous names and forms as the relative existence is dependent on the transformation of the *Brahma* without the loss of its' self or its' identity. The *Nirguna Brahma* thus, on the accountability of other aspects such as outer existence, and inner reality such as *S'abda* form the *Saguna Brahma*. *S'abda* is a representation of the bigger construct collapsed into a nutshell. The well-defined capacity of this highest reality is perceived as *Brahma,* which is then the representation of linguistic philosophy termed as *Śabda Brahma*. This motivation to seek or attain that eternal truth based on the semantic relations between the existence and the inference of it define the cognitive aspect of the *Brahma.* This together in conglomerate thus intends that speech is a form of experience and that this experience is derived from realisation and thought, and that thought is derived from the concept which is fluid in nature and has a quality that needs to be perceived with the help of grammar and parts of speech.

Thus, to seek that eternal truth in its' unadulterated and raw and pure form (which is *Nirguna* in nature) which is never changing and is fixed one needs to go through the tangible form (which is *Saguna* in nature). It is perceived by various schools of philosophy and grammatical traditions in various ways and aspects and hence it is devoid of a non-singular interpretation.

6. Bibliography and Referencing

Sharma, P. S. *THE KĀLASAMUDDEŚA OF BHATṚHARI'S VĀKYAPADĪYA.* Delhi: Motilal Banarsidass.

Varma, S. (1970). *VĀKYAPADĪYAM (BRAHMAKĀṆḌA) of SHRI BHATṚHARI.* New Delhi: Munshiram Manoharlal.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

An Acoustical and Neuro-cognitive Study on the Effects of Lyrics in Song from Non-linear Perspective

Archi Banerjee[a,b], *Shankha Sanyal*[b], *Souparno Roy*[b], *Priyadarshi Patnaik*[a] & *Dipak Ghosh*[b]

[a]IIT Kharagpur; [b]Jadavpur University, India

ARTICLE INFO

Article history:
Received 15/05/2020
Accepted 06/08/2020

Keywords:
Singing,
Humming,
Complex systems,
Non-linear analysis,
DFA

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Over thousands of years the language of communication between mankind has evolved and following the path gradually melody and words came into this world. A perfect marriage between the lyrics and melody gives birth to a "song". But due to the unavailability of proper scientific research and advanced analysis techniques, we are yet to know how the lyrics and melody controls our perception and cognition process while listening to a song. What happens when the lyric part of a song is separated from its melody and the melody of the song is simply hummed/ sung without any meaningful words? Would the emotional appraisal remain same with that of the song or will it change altogether? In the present work, the main aim is to quantify the acoustical contribution as well as the neuro-cognitive impacts of lyrics in different *bandishes* of Indian Classical (*Raga*) Music evoking contrast emotions. For this, recordings were taken from a professional female singer who was asked to consecutively sing (with proper meaningful lyrics) and hum (without using any lyric or meaningful words) four songs (two slow tempo *vilambit bandishes* and two faster tempo *drut bandishes*) of two different *ragas* depicting opposite/ contrast emotions: joy-sorrow, each of which were later put to analysis. Later to study the neuro-cognitive impact of the same, EEG signals were recorded for 5 musically untrained participants (who did not participate in the audio recording) with different sets of song-humming (i.e., with lyric-without lyric) versions of the same melodic content as stimuli. Both the acoustical and EEG time series data were analysed using the latest state of the art non-linear technique called DFA (Detrended Fluctuation Analysis). Outcomes reveal that the long range temporal correlations present in an acoustic signal depends both on the melodic and lyrical content of the renditions while acoustical and neuro-cognitive impact of lyrics within a song varies significantly with variation in tempo.

1. INTRODUCTION

Music is one of the oldest and fundamental mediums of communication between every species in the animal kingdom. With evolution over billions of years the nature and complexity of these communications and hence their music changed. Human beings invented another medium of communication: language – with meaningful words to express different facts and emotions. Then

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Archi Banerjee
Email: archibanerjee7@gmail.com

came the "song" which combines the words with the melody to create a complete musical experience. The musicians say "When a marriage happens between a lyric and a melody, only then a true song is born"! Elaborately speaking, a proper song should express the true meaning of its lyrics through the movements of its melodic structure. A song is 1) vocally produced, 2) linguistically meaningful and 3) melodic (Chow & Brown, 2018). Music is one of the most fundamental and strongest emotion evoking stimuli of this world but in the domain of musically induced human emotion perception, very little scientific facts are known till date. A question has been whirling in the musical fraternity for generations – "Which impacts more: Melody or lyrics?" – the answer to which is still unknown. It is well established that both lyrics and melody play distinctly different but important roles in portraying the emotional content of a particular song, though, during listening to music, the audience listens to the lyric and melody together and tends to get an idea about the mood of the whole song. But it is yet unknown if melody is transmitted faster than lyrics in the human brain and dominates the perceived emotional response. Also till date we are in dark about the basic scientific fact if lyrics and melody of a song are processed in the same brain regions or they demand completely different cognitive engagements. Few basic questions leading to this study are: What happens when the lyric part of a song is separated from its melody and the melody of the song is simply hummed/ sung without any meaningful words? Would the emotional appraisal remain the same with that of the song or will it change altogether? What would happen if the lyric part is separated from the melody in a vocal music and only the melody is conveyed to the audience as a separate entity altogether? Within a song, rhythm and pitch/ amplitude modulation act as the modes which combine the words together to convey the emotional content imbibed within. Perceptually, it is known that within the melodic structure of a song the presence of these pitch, amplitude modulation and rhythmic patterns brings out the true colours of the lyrical and melodic imagery but the mathematical analogue of the same is still unknown. In this study we envisaged to develop a scientific classification system taking into consideration songs (*Raaga Bandishes)* of Indian Classical Music (ICM) having decently strong/ emotion evoking lexical contents. The present work explores the acoustical contribution as well as the neuro-cognitive impacts of lyrics in different *bandishes* of ICM evoking contrast emotions. For this study recordings were taken from one female singer (receiving training in Indian classical music for 20 years and also a performer) who was asked to consecutively sing (with proper meaningful lyrics) and hum (without using any lyric or meaningful words) four songs (two slow tempo *vilambit bandishes* and two faster tempo *drut bandishes*) of two different *raagas* depicting opposite/ contrast emotions: joy-sorrow, each of which were later put to analysis. Subsequently EEG signals were recorded for 5 naïve i.e., musically untrained participants (who did not participate in the audio recording) with different sets of song-humming (i.e., with lyric-without lyric) versions of the same melodic content as stimuli to study their corresponding bio-signal (EEG) attributes.

For humans, there are other forms of auditory communication, like speech, but the difference is that music is more universal as the melody is free from language barrier. It has also been shown that not only humans, plants (Gadani & Mehta, 2002; Reddy & Raghavan, 2013) and animals (Uetake, Hurnik & Johnson, 1997) show positive responses under the effect of musical stimuli. Cognitive systems also underlie musical performance and sensibilities. That is why music intersects with cultural boundaries, facilitating our "social self" by linking our shared experiences and intentions. Music is linked to learning, and humans have a strong pedagogical predilection. Learning not only takes place in the development of direct musical skills, but in the connections between music and emotional experiences (Schulkin & Raglan, 2014). Results from

a wide range of investigations over the past century suggest that the various attributes of music, such as intensity (loudness), tempo, dissonance, and pitch height, are strongly associated with emotional expressions. In particular, changes in any of these attributes are correlated with changes in emotional interpretation (Ilie and Thompson 2006) and affective experience (Husain, Thompson, and Schellenberg, 2002; Ilie and Thompson 2011; Thompson et al. 2001). Such attributes contribute to an emotional code that may be employed by composers and performers to communicate emotions in music, or by speakers when they communicate emotions in their tone of voice (Juslin and Laukka 2003). One important cue is tempo. Melodies that are played at a slow tempo tend to evoke emotions with low energy such as sadness, whereas melodies that are played at a fast tempo tend to evoke emotions with high energy, such as anger or joy. To scientifically investigate the emotional significance of tempo, Hevner (1935) presented listeners with several pieces of classical music performed at slow (63–80 bpm) and fast (102–152 bpm) tempo.

EEG signals have been used by brain researchers and human computer experts to recognize human emotions (Subha et al., 2010). Shahabi et al. (Shahabi & Moghimi, 2016) proposed a non-invasive assessment tool for the automatic detection of musical emotions. They investigated how the brain is associated with joyful, melancholic, and neutral music. In (Bajaj and Pachori, 2014), four different human emotions are classified from EEG signals using wavelet based features. Three emotions are classified by Support Vector Machines (SVM) on the basis of time-frequency features (Chanel, Kierkels, Soleymani, & Pun, 2009). Soleymani et al (Soleymani et al., 2015) suggested a user-independent multimodal emotion recognition system using EEG technique, which classifies three emotional states in response to video clips. Petrantonakis et al (Petrantonakis and Hadjileontiadis, 2010) classified four different human emotions using four different machine learning techniques from brain signals using Higher Order Crossings (HOC) and Hybrid Adaptive Filtering (HAF). EEG signals are classified in response to music and a relationship between music and emotions is also established using brain maps (Lin et al., 2010). Hadjidimitriou et al (Hadjidimitriou & Hadjileontiadis, 2012) proposed an EEG-based music liking and disliking scheme using different time frequency analysis techniques including Hilbert Huang Spectrum (HHS) and Zhao Atlas Marks (ZAM). They compare results on the basis of different machine learning algorithms such as SVM and k-nearest neighbours (K-NN). In (Hadjidimitriou and Hadjileontiadis, 2013), music appraisal responses are classified using EEG signals based on topographical maps and time-frequency distributions. Daly et al. (Daly et al., 2014) presents some results in which subjects report their induced emotional responses and found neural correlation between beta and gamma bands based on EEG signals.

Previous knowledge suggests that music shows a very complex behaviour: at every instant components (in micro and macro scale: pitch, intensity, timbre, accent, duration, phrase, melody etc) are closely linked to each other. Similarly, EEG signals, coming out as the result of interaction between billions of neurons within the brain, feature a highly complex waveform. These properties are peculiar of systems with chaotic, self organized, and generally, non linear behaviour. Therefore, the analysis of music/EEG using linear and deterministic frameworks seems not to be sufficient. So, a non-deterministic/chaotic approach is needed in understanding the EEG/music signals. In this context fractal analysis of the signal which reveals the geometry embedded in the signal assumes significance. The Detrended Fluctuation Analysis (DFA) is a latest state of the art non-linear technique which essentially computes the Long range temporal correlations (LRTC) present in the audio/EEG signals and uses a scaling exponent (called Hurst Exponent) to quantify them (Banerjee et.al, 2016). In this study, DFA was employed to calculate

the Hurst Exponent values for all of the recorded auditory signals and subsequently the scaling exponent values were compared for each pair of audio signals (having the same melodic structure with/without meaningful words i.e., song-humming) to understand the acoustic contribution of the lyrics in a song from a quantitative approach. The same DFA technique was used also to analyze the EEG signals recorded when the participants listened to the song-humming versions of the same melody to understand the impact of lyrics on the human brain for different genres of song with varying emotional and lyrical content. This is a pilot study in the context of Indian music which endeavours to analyze the contribution of lyrics in songs of different genres as well as different emotional content in both acoustic and neuro-cognitive domains.

2. EXPERIMENTAL DETAILS

Choice of audio signals:

One female singer (receiving training in Indian classical music for 20 years and also a performer) was asked to consecutively sing (with proper meaningful lyrics) and hum (without using any lyric or meaningful words) four *bandishes* (songs) having strong lexical content structured on two different *ragas* (traditionally depicting opposite/ contrast emotions: joy-sorrow) of Indian Classical Music. The two *ragas* chosen by the singer are *Multani* (Sad) and *Hamsadhwani* (Happy). Total 8 clips were recorded (Song and humming versions of each of the following 4 *bandishes): i) Multani vilambit, ii) Hamsadhwani vilambit, iii) Multani drut and iv) Hamsadhwani drut.* Average tempo of the *vilambit bandishes* were within 30-40 beats per minute whereas the *drut bandishes* were sung at an average tempo of 90-120 beats per minute.

Acquisition of EEG signals:

Subjects chosen for EEG:

5 right handed adults (3 male and 2 female) voluntarily participated in this study. None of them had any conventional musical training in Indian classical music. Neither reported any neurological disorders or auditory impairment. Their ages were between 18 to 28 years (SD=2.25 years). All experiments were performed at the Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata. Informed consent was obtained from each subject according to the ethical guidelines of the Ethical Committee of Jadavpur University.

Experimental Protocol for EEG:

During the EEG acquisition period, the 5 subjects were subjected to 8 pairs of audio clips with each pair consisting of a humming and a song version of any of the paragraphs of the recorded 4 songs. First, the baseline (that is, an eyes closed resting condition) was recorded for each subject before the start of the experiment with 2 minutes of 'no stimuli' condition. Then the audio clips were played using the computer-sound system (Logitech R _ Z-4 speakers, placed 120 cm in front of the head of the subjects) with very low S/N ratio keeping the volume fixed throughout the experiment. Each subject was prepared with an EEG recording cap with 19 electrodes (Ag/AgCl sintered ring electrodes) placed in the international 10/20 system. **Figure 1** depicts the positions of the electrodes. Impedances were checked below 50 kΩ. The EEG recording system (Recorders and Medicare Systems) was operated at 256 samples/s recording on customized software of RMS. The data was band-pass-filtered between 0 and 50 Hz to remove DC drifts. Each subject was seated comfortably in a relaxed condition in a chair in a shielded measurement cabin. They were also asked to close their eyes. After initialization, an approx. 20 min recording period was started with 4 sets of audio clips (where each set maintained the following playing

order – i) *sthayi* without lyric, ii) *sthayi* with lyric, iii) *antara* without lyric, iv) *antara* with lyric; each of the 4 parts separated by a gap of 20 seconds) and the following protocol was followed:

1. 2 minutes Before Music (Resting Condition)
2. Set 1 (*Raga Multani vilambit bandish)*
3. 30 seconds No Music
4. Set 2 (*Raga Hamsadhwani vilambit bandish)*
5. 30 seconds No Music
6. Set 3 (*Raga Multani drut bandish)*
7. 30 seconds No Music
8. Set 4 (*Raga Hamsadhwani drut bandish)*
9. 2 minutes After Music (Resting Condition)

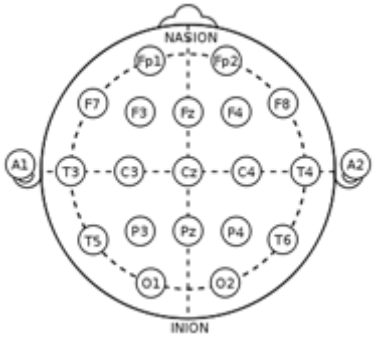


Fig. 1: The position of electrodes according to the 10-20 international system

3. METHODOLOGY

Processing and analysis of audio signals:

All recorded audio (Song and humming) signals were normalized to 0 dB. Each of these sound signals was digitized at the sampling rate of 44.1 KHz, 16 bit resolution and in a mono channel. Each of the 8 audio signals was then divided into 2 parts according to the *sthayi* and *antara* part of the lyrics - which are respectively the first and second paragraph of the song chosen. Finally, 8 pairs of music clips were prepared with song and hum versions of the same melodic content. Scaling exponent was calculated for each part with the help of DFA technique. Scaling exponent values were compared for the parts having the exact same melodic content (in the singing as well as in the humming versions i.e., with and without lyrics).

Processing and analysis of EEG signals:

First the complete 20 minute raw EEG of a single subject was cut along the temporal markers for the different experimental conditions. Noise cleaned EEG data were obtained for all the electrodes using the EMD technique and then the noise cleaned data was used for further analysis (DFA) to study the different kinds of acoustic stimuli induced EEG features. We identified four different lobes of the brain namely frontal, parietal, temporal and occipital whose functions tally with this work and the electrodes chosen for analysis are F3, F4, F7, F8, Fz (Frontal lobe); P3, P4 (Parietal lobe); O1, O2 (Occipital lobe); T3, T4 (Temporal lobe). Hurst/Scaling exponent was calculated for each part using DFA technique. The same procedure is followed for each of the 5 subjects. The average changes in the scaling exponent values between each pair of experimental conditions (same melodic content without and with lyrics) are plotted.

DFA has been developed to quantify the symmetry scaling behaviour or the correlation properties present in non-stationary signals e.g., in physiological time series, because long-range correlations can also come from the artifacts of the time series data. In our study, DFA technique has been applied to quantify the scaling behaviour of the fluctuations in each of the pre-processed audio signal parts and their EEG counterparts and finally the scaling exponents were compared to study the acoustical and neuro-cognitive contribution of each pair of humming and song signals keeping the melody exactly same.

Detrended Fluctuation Analysis

To compute the Hurst exponent of a time series: $x_1, x_2, \ldots, x_N$ using Detrended Fluctuation Analysis or DFA technique [13], firstly integration of x is done to form a new series $y = y_1, \ldots, y_N$ where

$$y(k) = \sum_{i=1}^{k} (x_i - \bar{x}) \quad \ldots\ldots\ldots (1)$$

$\bar{x}$ is the mean of $x_1, x_2, \ldots, x_N$.

The integrated series is then sliced into boxes of equal length or intervals of size n. In each box of length n, a least-squares line is fit to the data. The least square line represents the trend in that box. The coordinates of the straight line segments are denoted by $y_n(k)$. The root-mean-square fluctuation of the integrated series is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(k) - y_n(k))^2} \quad \ldots\ldots\ldots (2)$$

where the part $y(k) - y_n(k)$ is called detrending. The relationship between the detrended series and interval lengths can be expressed as

$$F(n) \propto n^{\alpha} \quad \ldots\ldots\ldots (3)$$

where α is expressed as the slope of a double logarithmic plot log $F(n)$ versus $log(n)$ (as shown in representative **Fig. 2 (a-c)** . The parameter α (scaling exponent, autocorrelation exponent, self-similarity parameter etc.) represents the auto-correlation properties of the signal.

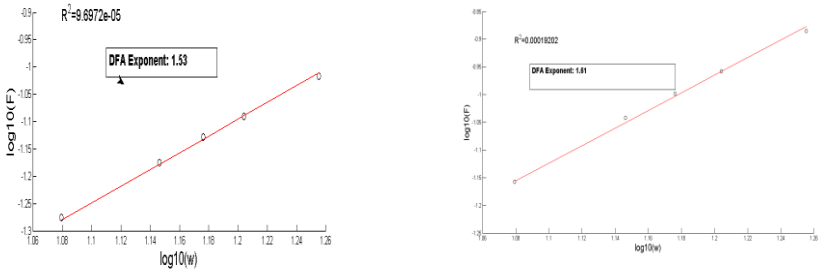


Fig. 2 (a, b): DFA scaling exponent from double logarithmic plots for a) humming and b) song

DFA technique was applied following the NBT algorithm used by Hardstone et.al (2012). The scaling exponent provides a quantitative measure of long range temporal correlation (LRTC) that exists in the audio as well as EEG signals. When the auditory or EEG waveform is completely uncorrelated (Gaussian or non-Gaussian probability distribution), the calculation of the scaling exponent results 0.5, also called white noise. When applied to audio/EEG data with LRTC, power-law behaviour will generate scaling exponents with greater than 0.5 and less than 1. As the scaling exponent increases from 0.5 to 1, the LRTC in the audio/EEG is more persistent (decaying more slowly with time). If a scaling exponent is greater than 1, the LRTC no longer exhibits power law behaviour. Finally, if the scaling exponent = 1.5, this indicates Brownian noise, which is the integration of white noise.

Significance of Scaling Exponent from DFA/AFA	
Exponent	Remarks
<0.5	anti-correlated signal
=0.5	uncorrelated signal (white noise)
>0.5	positive autocorrelation in the signal
=1	1/f noise
=1.5	Brownian noise or random walk

4. RESULTS & DISCUSSION

DFA technique was applied to calculate the Hurst or Scaling exponent for each part of the recorded audio signals and then the Scaling exponent values were compared for the parts having the exact same melodic content (in the singing as well as in the humming versions i.e., with and without lyrics) in attempt to quantify the acoustical contribution of lyrics in a song (*bandish*) in Indian classical music. The following bar graphs (**Fig. 3a & 3b)** represent the variation of DFA scaling exponent in the song and humming (i.e., with and without lyrics) versions of the chosen 4 *bandishes* (in Raga *Hamsadhwani* and *Multani* both in *vilambit* and *drut laya*).

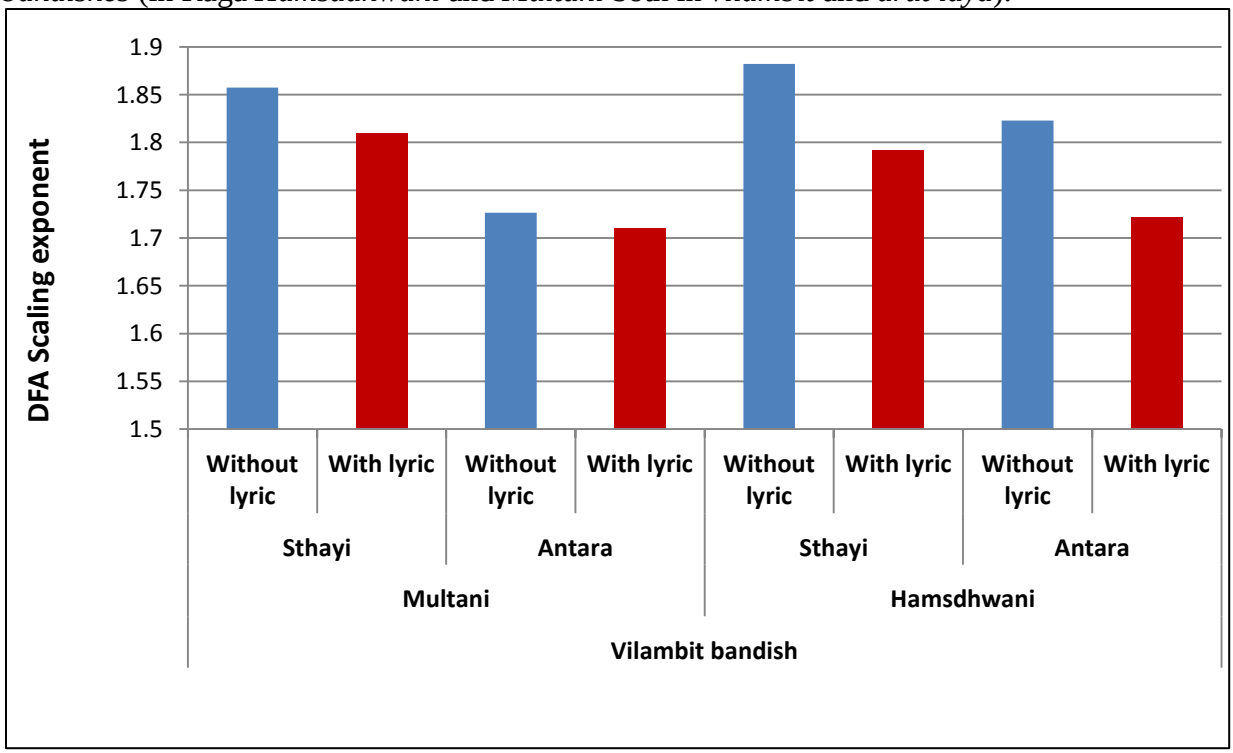


Fig. 3a: Variation in DFA scaling exponent in acoustical signals of 2 *bandishes* with and without lyrics in *Vilambit laya*

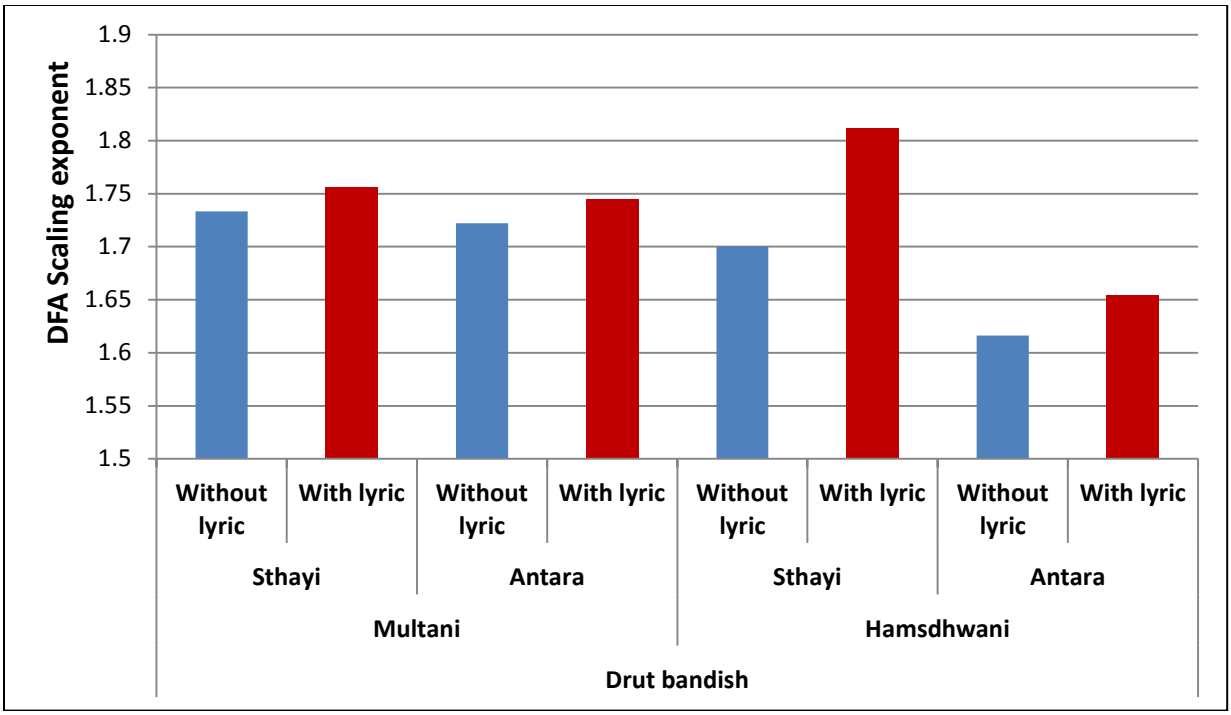


Fig. 3b: Variation in DFA scaling exponent in acoustical signals of 2 *bandishes* with and without lyrics in *Drut laya*

Acoustic signal analysis **(Fig. 3a & 3b)** reveals that

- For all 4 *bandishes*, i.e., *Hamsadhwani* and *Multani vilambit* and *drut bandishes*, average DFA scaling exponent values for *antara* parts are lower than the corresponding *sthayi* parts. This trend is replicated both in the song and humming version of the same melodic content. This may be attributed to the difference in overall average pitch difference between the melodic nature of *sthayi* and *antara* in a song (*bandish* in our case). Usually *antara* part in a song features higher pitch range on an average compared to the *sthayi* part, which can contribute to the lowering of acoustical signal complexity.
- In case of *vilambit bandishes* (**Fig. 3a)**, the song (with lyric) versions of same melodic content feature lower scaling exponent values than the humming (without lyric) versions in both *sthayi* and *antara* parts of the two ragas. But, this trend is just reverse in case of *drut bandishes* (**Fig. 3b)** where the song versions yield higher scaling exponent values than the humming counterparts. Several reasons may contribute together to result in this. In an attempt to analyse the probable causes we think that probably, in *vilambit laya* (low tempo) the compactness of lyric within a *bandish* is lower than that in the *drut laya* (higher tempo) *bandishes*. So, in the lower tempo range acoustical signal complexity is dominated by the contribution of the melodic structure rather than the lyrical contents. In other words, because of the lower tempo of the *vilambit bandishes*, in some cases, the scattered nature of the lyrics in *vilambit laya* can interfere with the richness of the melodic structure of the song. In case of *drut bandishes* the lyric usually sounds more compact because of the higher tempo range and the *bandish* lyric behaves like that of a composed song. The lyric adds its own rhythm to the already existing melodic rhythm and plays a significant contribution in the acoustical signal complexity in addition to the melodic structure.

- The rate of decrement in scaling exponent values from humming to song in *vilambit bandish* parts is higher than rate of increment of the same in *drut bandish* parts. This can be attributed to the dominance of *Taal* or beat patterns present in case of *drut bandishes*. In higher tempo the constant presence of beat patterns sometimes suppresses the additional impact of lyrical contribution over melodic structure.
- On an average, *vilambit bandishes* yield higher scaling exponent values than *drut bandishes*. Probably the tempo of the song is another contributing factor in determination of the acoustical signal complexity and higher tempo ultimately leads to decrement in scaling exponent values.
- In *vilambit laya, Hamsadhwani bandish* features higher scaling exponent than *Multani bandish* but the trend is reverse for *drut bandishes* of the two *ragas*. So, apparently, from these results we failed to categorize the individual impact of lyric in songs of different emotions. But, to verify this result, recordings of a large number of *bandishes* evoking different emotions should be taken from various artists and analysed using latest state of the art techniques.

To study the brain response (EEG) of each participant corresponding to the above mentioned music stimuli, Hurst/Scaling exponent was calculated for the chosen 11 electrodes (F3, F4, F7, F8, Fz in Frontal lobe; P3, P4 from Parietal lobe; O1, O2 from Occipital lobe; T3, T4 from Temporal lobe) for each experimental condition using DFA technique. The average changes in the scaling exponent values between each pair of humming-song version (same melodic content without and with lyrics) are plotted in the following bar graphs (**Fig. 4 a-e).** For every pair of clips in the following graphs the even number of clips represent the song version (with lyrics) and the odd number of clips represent the humming version (without lyrics) of the same melodic structure or the same part of the *bandishes*.

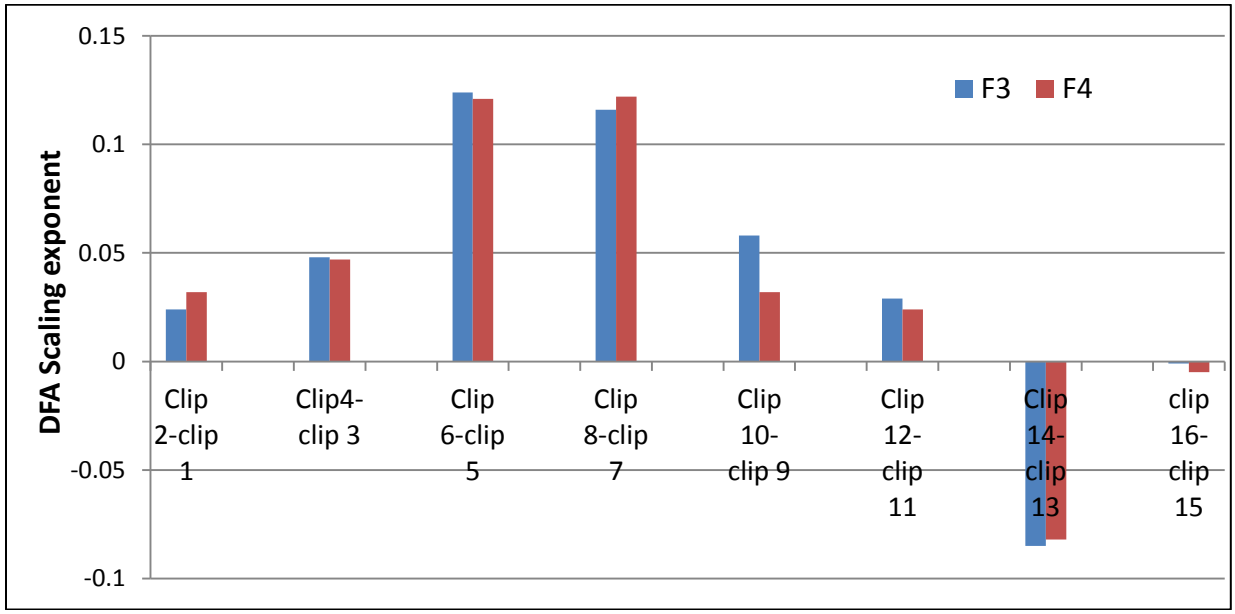


Fig. 4a: Differences in DFA scaling exponent in Frontal electrodes (F3, F4) for different song-humming pairs

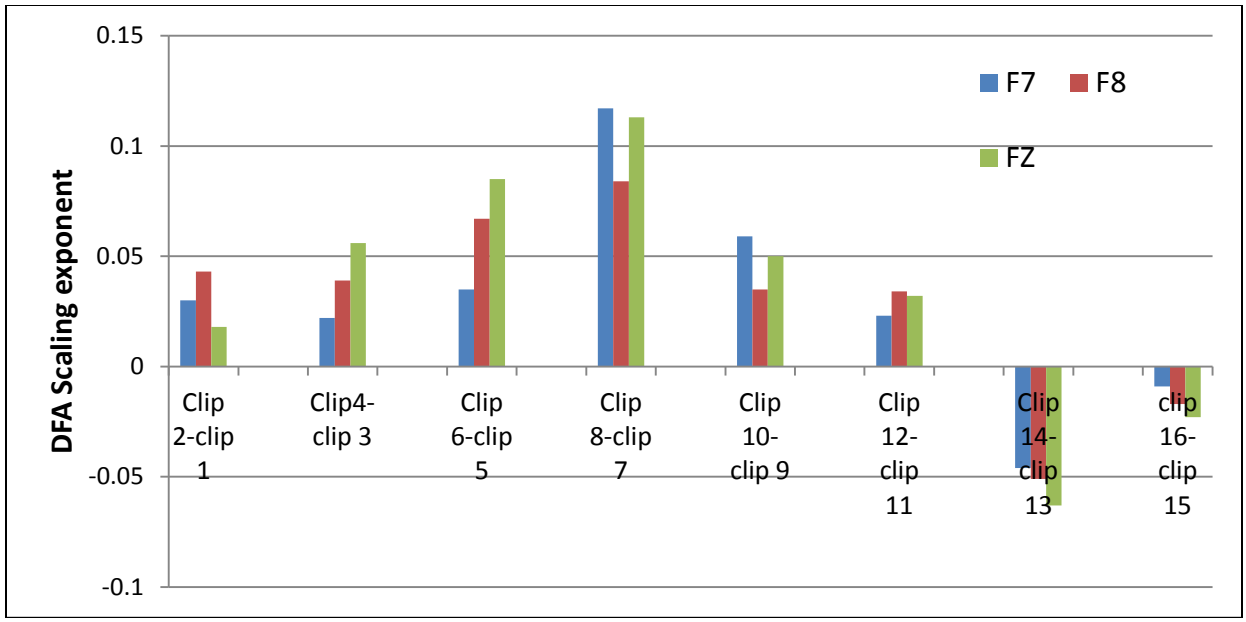


Fig. 4b: Differences in DFA scaling exponent in Frontal electrodes (F7, F8 & FZ) for different song-humming pairs

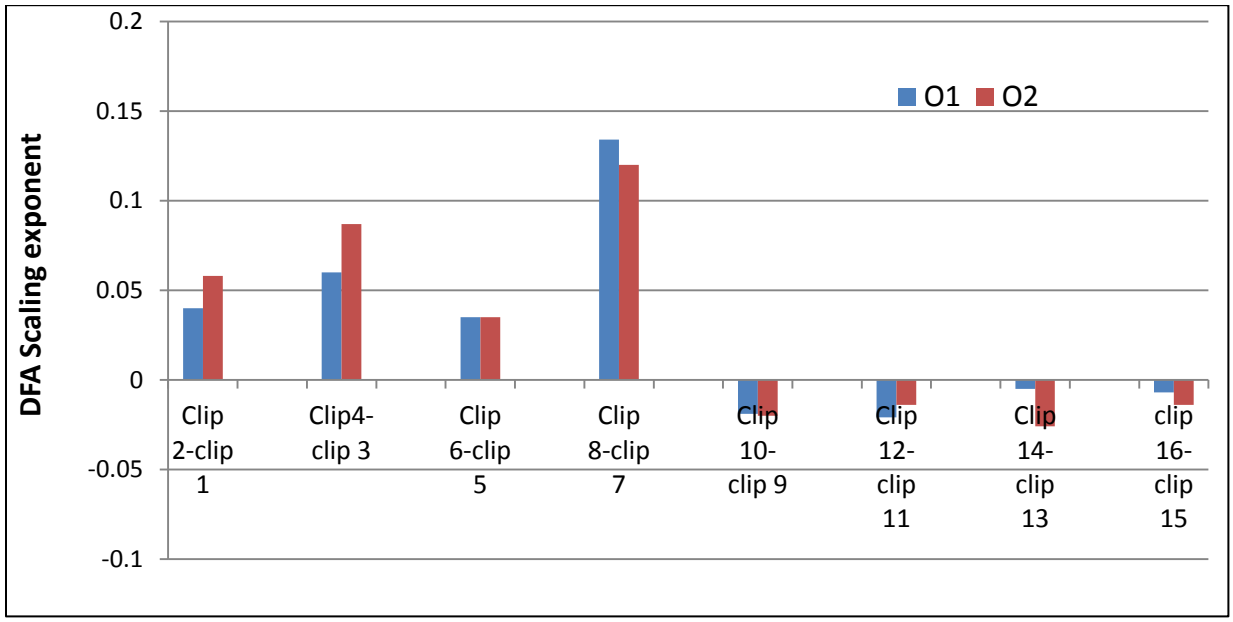


Fig. 4c: Differences in DFA scaling exponent in Occipital electrodes (O1 & O2) for different song-humming pairs

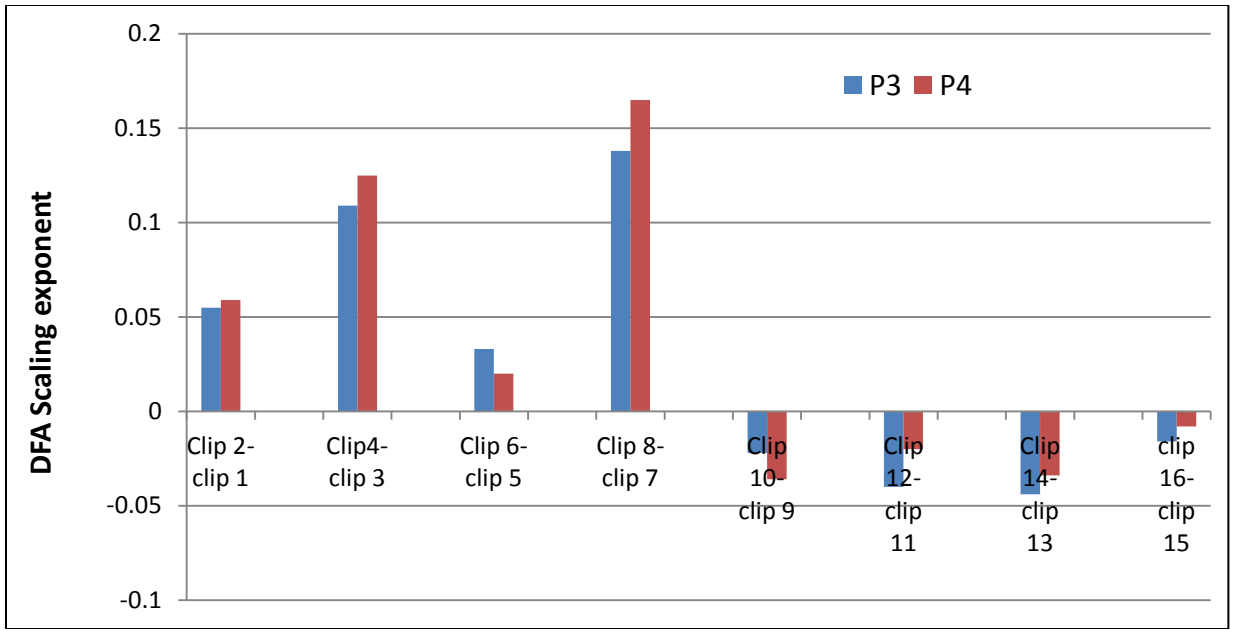


Fig. 4d: Differences in DFA scaling exponent in Parietal electrodes (P3 & P4) for different song-humming pairs

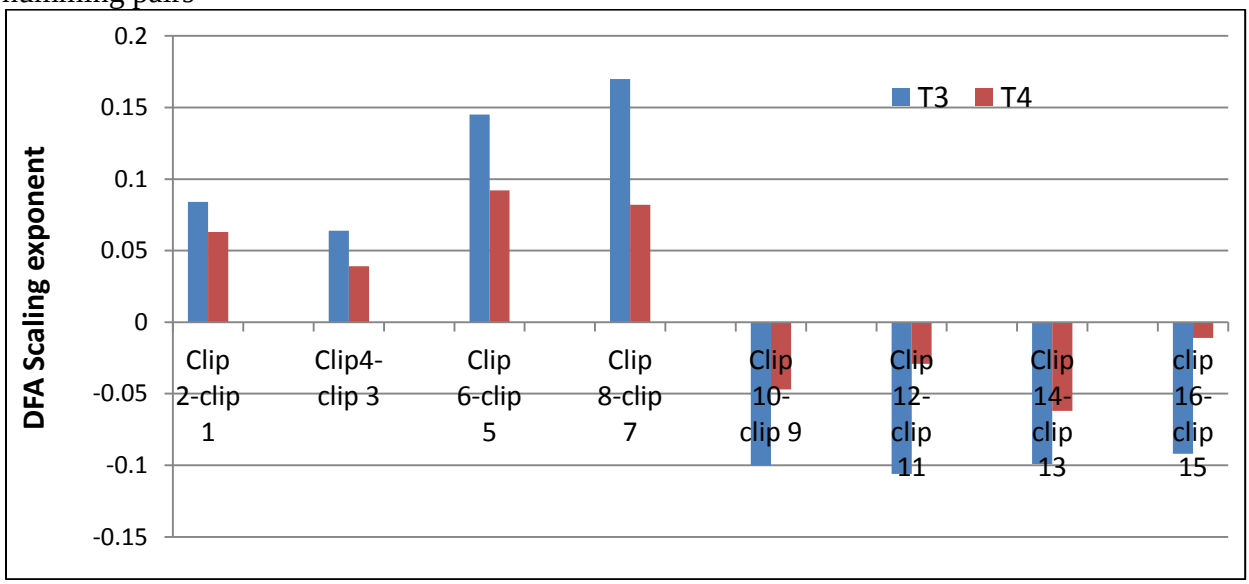


Fig. 4e: Differences in DFA scaling exponent in Temporal electrodes (T3 & T4) for different song-humming pairs

EEG signal analysis (**Fig. 4 a-e)** reveals that

- ❑ Consistently for all frontal electrodes: F3, F4, F7, F8 & FZ (**Fig. 4a & 4b)**, average DFA scaling exponent values for song (with lyrics) versions of the same melodic content are higher than the corresponding humming (without lyrics) versions except for both the *sthayi* and *antara* part of the *Hamsadhwani drut bandish.*
- ❑ Among all 4 chosen *bandishes*, *Hamsadhwani vilambit* yield maximum change in scaling exponent from humming to song for all chosen frontal electrodes. This trend is similar across all other lobes of brain considered for our analysis.
- ❑ In Occipital (O1 & O2), Parietal (P3 & P4) and Temporal (T3 & T4) electrodes, the song versions of *vilambit bandishes* feature higher scaling exponent values than the humming

versions in both *sthayi* and *antara* parts of the chosen two *ragas*, whereas, in case of *drut bandishes*, the song versions yield lower scaling exponent values than the humming counterparts.

- In the Frontal, Occipital and Parietal electrodes, the average changes in scaling exponent for *drut bandishes* are significantly lower than the *vilambit bandishes*. In Temporal lobe, though, the average changes of scaling exponent for both *vilambit* and *drut bandishes* are almost equivalent.

These observations may shade some light on the cognition process of a song lyric in our brain, particularly in the scientifically much unexplored domain of Indian Classical Music. Probably, during cognition of lyrics within a song, tempo plays a key role which determines the changes in brain signal complexity. During conscious listening, in *vilambit laya* (low tempo) the emotion evoking contribution of lyric within a *bandish* is higher than that in the *drut laya* (higher tempo). The dominance of beat pattern in *drut bandishes,* in some cases if the lyric is not sung with proper scansion and distribution over time, may interfere with the lyrical meaning and imagery of the song and resist them from getting conveyed to the audience very clearly.

5. **CONCLUSIONS**

Combining the results from acoustic analysis and human brain response (EEG) analysis using the same technique, we find some interesting results.

- The DFA scaling exponent (which is a manifestation of the long range temporal correlations present in the time series) of an acoustic signal depends both on the melodic and lyrical content of the renditions. The change in self similarity of the signals can be attributed to the contribution of tempo, rhythm, amplitude and pitch modulation along with the lexical content.
- The acoustical and neuro-cognitive impact of lyrics within a song varies significantly with variation in tempo, but we could not identify any such distinction for variation in emotion (Happy-Sad in our case).
- In this study, we attempt to quantify the impact of lyrics within a song in both acoustical and neuro-cognitive level with the help of a unique scaling exponent called the Hurst exponent which is a novel step in the domain of music research from a hardcore scientific perspective.

Potentials of this study:

- Analysis with a greater number of audio samples as well as EEG participants will lead to more accurate understanding of the individual role of lyrics and melody in a song. This study is a precursor in that direction.
- The playing order of the recorded audio signals (humming-song pairs) may have affected the cognitive impact of the lyrics in different participants. To determine the individual contribution of lyrics the experiment must be repeated by randomly shuffling the playing order of the samples.
- In Indian Classical Music, the concept of "absolute happy" and "absolute sad" music is not entirely true; i.e., a particular *raga* can portray an amalgamation of a number of perceived emotions. So, a study comparing the results of this work with a statistically significant large data pool of human response needs to be performed to draw anything more conclusive.

6. ACKNOWLEDGEMENTS

One of the authors, AB acknowledges the Department of Science and Technology (DST), Govt. of India for providing (SR/CSRI/PDF-34/2018) the DST CSRI Post Doctoral Fellowship to pursue this research work. SS acknowledges the JU RUSA 2.0 Post Doctoral Fellowship (R-11/557/19) and Acoustical Society of America (ASA) to pursue this research.

7. REFERENCES:

Banerjee, A., Sanyal, S., Patranabis, A., Banerjee, K., Guhathakurta, T., Sengupta, R., Ghosh, D. and Ghose, P., 2016. Study on brain dynamics by non linear analysis of music induced EEG signals. Physica A: Statistical Mechanics and its Applications, 444, pp.110-120.

Chanel, G., Kierkels, J. J., Soleymani, M., & Pun, T. (2009). Short-term emotion assessment in a recall paradigm. International Journal of Human-Computer Studies, 67(8), 607-627.

Chow, I., & Brown, S. (2018). A Musical Approach to Speech Melody. *Frontiers in psychology*, *9*, 247.

Daly, I., Malik, A., Hwang, F., Roesch, E., Weaver, J., Kirke, A., ... & Nasuto, S. J. (2014). Neural correlates of emotional responses to music: an EEG study. Neuroscience letters, 573, 52-57.

Gadani, M., & Mehta, D. (2002). EFFECT OF MUSIC ON PLANT GROWTH. Plant Sciences, 8(2), 253-259.

Hadjidimitriou, S. K., & Hadjileontiadis, L. J. (2012). Toward an EEG-based recognition of music liking using time-frequency analysis. IEEE Transactions on Biomedical Engineering, 59(12), 3498-3510.

Hadjidimitriou, S. K., & Hadjileontiadis, L. J. (2013). EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings. IEEE Transactions on Affective Computing, 99(1), 1.

Hardstone, R., Poil, S. S., Schiavone, G., Jansen, R., Nikulin, V. V., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2012). Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Frontiers in physiology*, *3*, 450.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. Music Perception: An Interdisciplinary Journal, 20(2), 151-171.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. Music Perception: An Interdisciplinary Journal, 23(4), 319-330.

Juslin, P. N., & Laukka, P. (2003). Emotional Expression in Speech and Music: Evidence of Cross-Modal Similarities. Annals of the New York Academy of Sciences, 1000(1), 279-282.

Liu, J., Zhang, C., & Zheng, C. (2010). EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters. Biomedical Signal Processing and Control, 5(2), 124-130.

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. IEEE Transactions on Information Technology in Biomedicine, 14(2), 186-197.

Reddy, K. G., & Ragavan, R. (2013). Classical ragas: A new protein supplement in plants. Indian Journal of Life Sciences, 3(1), 97.

Schulkin, J., & Raglan, G. B. (2014). The evolution of music and human social capability. Frontiers in neuroscience, 8, 292.
Shahabi, H., & Moghimi, S. (2016). Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity. Computers in Human Behavior, 58, 231-239.
Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2015). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, *7*(1), 17-28.
Subha, D. P., Joseph, P. K., Acharya, R., & Lim, C. M. (2010). EEG signal analysis: a survey. Journal of medical systems, 34(2), 195-212.
Uetake, K., Hurnik, J. F., & Johnson, L. (1997). Effect of music on voluntary approach of dairy cows to an automatic milking system. Applied animal behaviour science, 53(3), 175-182.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Speech Rhythm in Malayalam Speaking Children with Hearing Impairment

Yeshoda Krishna, Revathi Raveendran, Sreeraj Konadath
All India Institute of Speech and Hearing, India

ARTICLE INFO
Article history:
Received 15/05/2020
Accepted 06/08/2020
Keywords:
Speech rhythm,
Malayalam,
Hearing Impairment,
Suprasegmental,
nPVI,
rPVI
Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh
Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU
Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT
The hearing sensitivity of an individual can affect the prosody because it provides necessary feedback for control over the speech. Auditory feedback affects both moment-to-moment and later control of speech, which also plays a major role in the acquisition of language; initially with lexical stress, grammatical categories and even emotional affect. The intervention strategies taken up during the rehabilitation of the children with hearing impairment primarily focuses on the segmental features of speech, even when it is realized that suprasegmental errors also have to be addressed to improve the speech intelligibility. Thus, knowledge about speech rhythm is crucial to provide this clinical population with holistic intervention strategies and helping the clinical population achieve naturalness of speech. Participants were 30 children (16 males and 14 females) with normal hearing (CWNH) in the chronological age range of 3.1 to 6.11 years and 30 children (18 males and 11 females) with severe to profound sensory neural hearing impairment (CWHI), using programmable digital behind the ear hearing aids with a language age of 3.1-6.11 years. The children were evaluated using a standardized test the Malayalam- Receptive and Expressive Language Tool (M-RELT) to have a language age within 3 to 7 years. Aided pure tone and speech audiometry were done to confirm the within speech spectrum thresholds.

1. Introduction

Prosody is the melody and rhythm of spoken language. Operationally, prosody can be defined as 'the suprasegmental features of speech that are conveyed by the parameters of fundamental frequency, intensity, and duration'; such suprasegmental features include stress, intonation, tone, and duration (Kent & Kim, 2008). One of the major components of prosody, rhythm is considered as the temporal patterning of speech events at the level of phonetic segments. The word rhythm indicates the perceptual distinction between the stressed and unstressed syllables,

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***
Corresponding Author: Revathi Raveendran
Email: revathirslp@gmail.com

which characterizes each language with its contrastive rhythm. Contrastive rhythm of a particular language imparts an aesthetic appeal and naturalness to the speech.
Low, Grabe, and Nolan (2000) developed normalized Pairwise Variability Index (nPVI) as well as raw Pairwise Variability Index (rPVI) which were able to successfully classify different languages into different rhythm categories in terms of their values either as low or high. nPVI measured the rhythm of vocalic segments and rPVI measured the rhythm of intervocalic segments. A PVI above 50 was considered as high and a PVI below 50 was considered as low (Merin, 2016). The participants of the current study were native Malayalam speakers. Malayalam is a syllable-timed language (Savithri, Jayaram, Kedarnath&Goswami, 2006).

The hearing sensitivity of an individual can affect the prosody because it provides necessary feedback for control over the speech. Auditory feedback affects both moment-to-moment and later control of speech, which also plays a major role in the acquisition of language; initially with lexical stress, grammatical categories and even emotional affect (Kuel, 2000; Yoshinago-Itano, 2000). The speech of children with hearing impairment, is characterized by errors in timing, voice quality and intonation(Osberger,1978; Rosenhouse, 1986). Most often, their speech will be filled with prolonged and/or inappropriate pauses, inaccurate stressed/unstressed syllable prolongations and imprecise transitions between words (Savithri, Johnsirani&Ruchi, 2008).

Savithri et al (2008) found that in native Kannada speaking children with hearing impairment in the age range of 5-10 years had longer intervocalic and vocalic durations in comparison with their age-matched typically developing children. The values were high in nPVI and low in rPVI, the rhythm of both groups were noted as unclassified. The results revealed that children with and without hearing loss used a much simpler syllabic structure in comparison to the adult speech structure. Merin and Savithri (2016) did a study on Malayalam speaking children with hearing impairment in the age range of 3-4 years. Though Malayalam was classified as a syllable-timed language by Savithri et al (2007) on adult native speakers of Malayalam, the speech rhythm in both children with and without hearing impairment was classified as mora-timed language. The children with hearing impairment were found to have an nPVI value of 0.223 and rPVI value of 0.293, while children with normal hearing sensitivity had nPVI of 0.184 and rPVI of 0.286. The authors specified that the results emphasized the need to include rhythm training in speech intervention strategies.

2. **Purpose:**

The intervention strategies in the rehabilitation of the children with hearing impairment primarily focuses on corrections of the segmental features of speech, knowing very well that the suprasegmental errors also have to be addressed to improve the speech intelligibility in these children. Thus, knowledge about speech rhythm is crucial to provide this clinical population with holistic intervention strategies and helping the clinical population achieve naturalness of speech.

Also, the contrastive rhythm being specific to each language, knowledge about the rhythm class of each language is desirable for framing accurate intervention goals.

3. **Aim:**

To study the speech rhythm of the children with hearing impairment who were the native speakers of Malayalam language.
Objective: To investigate and compare the speech rhythm of the language age matched children with hearing impairment and children with normal hearing.

4. **Method:**

Participants were 30 children (16 males and 14 females) with normal hearing (CWNH) in the chronological age range of 3.1 to 6.11 years and 30 children (18 males and 11 females) with severe to profound sensory neural hearing impairment (CWHI), using programmable digital behind the ear hearing aids with a language age of 3.1-6.11 years. The children of both the groups were evaluated using a standardized test the Malayalam- Receptive and Expressive Language Tool (M-RELT) to have a language age within 3 to 7 years. The CWHI underwent aided pure tone and speech audiometry tests to confirm the hearing acuity within speech spectrum thresholds using binaural hearing aids.

The authors formulated 10 concrete Malayalam sentences and their hand-drawn picture representations. These sentences were audio recorded into the CSL software (Kay Pentax, NJ) in a sound-treated room by the second author (female native speaker of Malayalam) using a unidirectional microphone maintaining mouth-microphone distance of six inches. The pictures were validated after checking their representability by testing on three typically developing Malayalam speaking children in the age range of 3.0-6.3 years.

Procedure: The participants were seated in an audiometric room. The stimuli were presented through calibrated loudspeakers at 50 dB HL at a distance of 1 meter and the azimuth of 0 degrees, routed through an audiometer along with the presentation of picture cards of each sentence, to facilitate the auditory stimulus. The participants were instructed to carefully listen to the auditory stimulus and then repeat the sentence.

Analysis: The responsesof the participants were analyzed using Praat software (Version: 5.2.01; Boersma &Weenink, 2009). The vocalic and intervocalic duration for each sentence was measured and tabulated for each participant. Normalised Pairwise Variability Index-vocalic (nPVI-V) and Raw Pairwise Variability Index-consonantal (rPVI-C) scores for duration were calculated using the formulae given by Low, Grabe, and Nolan (2000):

$$nPVI = 100 \times \frac{\left(\sum_{k=1}^{m-1} \left| (d_k - d_{k+1})/((d_k + d_{k+1})/2) \right|\right)}{(m-1)}$$

$$rPVI = \left(\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right|\right) / (m-1)$$

Statistical Analysis: The collected data was fed into the SPSS statistics software version 20. Descriptive Statistical analyses was done for nPVI – vocalic and rPVI – consonant values to find out the mean, median and standard deviation, to determine the rhythm class of the CWNH and CWHI. Inferential statistics was used to find the significant difference between the groups.

5. **Result& Discussion:**

The mean, median and standard deviation of nPVI-V and rPVI-C was calculated for each age group of both CWNH and CWHI and is listed in tables 1 & 2. The comparison graphs representing the differences between the nPVI and rPVI values are given in figure 1 & 2.The results of the Mann Whitney test showed that there was an overall significant difference between the PVI scores of the CWNH and CWHI groups ($Z = 2.15$; $p < 0.05$).

Table 1.*The mean, median and standard deviation of nPVI and rPVI values for the control group CWNH across the age group.*

		Age groups				
Groups	PVI values		3.1-3.11	4.0-4.11	5.0-5.11	6.0-6.11
		Mean	38.89	59.99	40.27	49.2
		Median	39.46	59.14	40.36	49.27
	nPVI	S.D	1.92	1.93	1.93	2.06
		Mean	100.23	89.11	59.68	80.23
		Median	99.65	89.27	60.41	80.13
CWNH	rPVI	S.D	1.45	1.56	1.56	2.15

Table 2.*The mean, median and standard deviation of nPVI and rPVI values for the control group CWHI across the age group.*

		Age Groups				
Group	PVI values		3.1-3.11	4.0-4.11	5.0-5.11	6.0-6.11
		Mean	52.92	64.66	57.01	38.74
		Median	53.01	65.24	57.32	38.46
	nPVI	S.D	2.02	1.54	1.66	2.03
		Mean	111.45	140.73	177.11	75.63
		Median	112.46	141.75	176.82	74.47
CWHI	rPVI	S.D	3.67	2.97	4.36	4.97

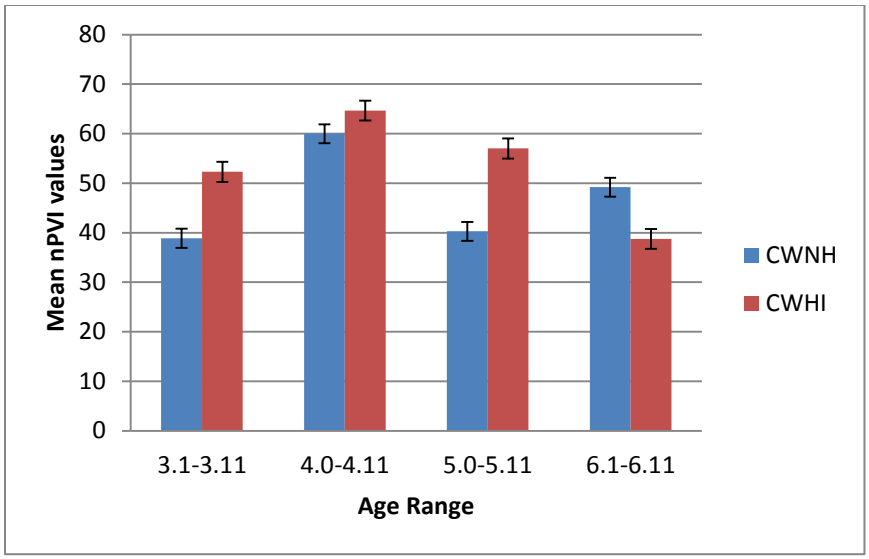


Figure 1: Comparison between the mean nPVI values of CWNH and CWHI groups.

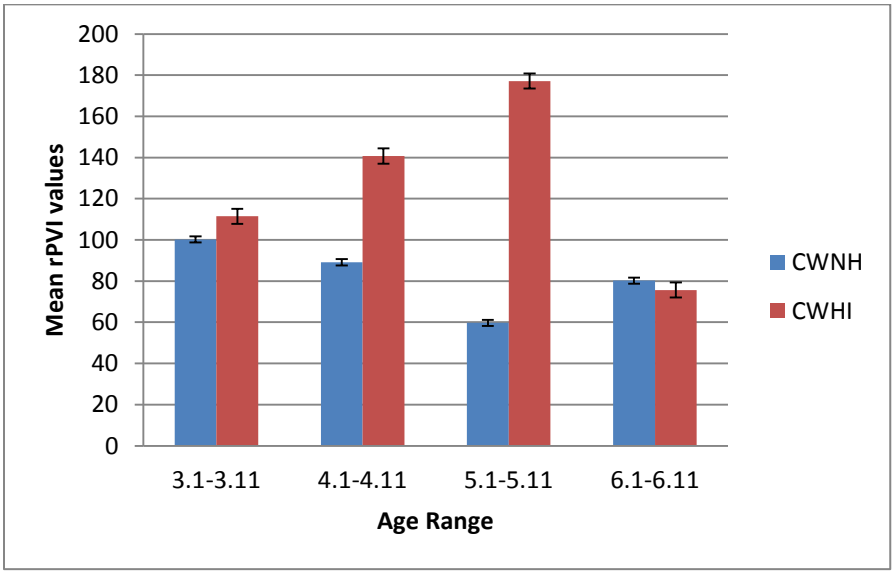


Figure 2: Comparison between the mean rPVI values of CWNH and CWHI groups.

Both nPVI and rPVI values were observed to be higher in CWHI when compared to CWNH, except for the 5.11-6.11 age groups. Higher nPVI and rPVI values could indicate that there were greater variations in the vocalic as well as the intervocalic segments. The speech of the children with hearing impairment were reported to have inconsistent prolonged vowels (Calvert,1962; Parkhurst & Levitt, 1978) as well as greater stressed syllables(Rosenhouse, 1986), which might have contributed to the high PVI values. The lack of accurate auditory feedback and auditory memory could have led to this deviant rhythm pattern in CWHI. The increase in the PVI values also indicates that the speech of CWHI is characterized with greater variations contributing to the

reduction of speech intelligibility. Also, the speech rhythm class of the CWHI group was observed to be unclassified.

There have been studies suggesting that children with hearing CIs or HAs are able to discriminate between different rhythmic patterns (Innes-Brown, 2013). Yet, the attention given to rhythm perception or production assessment and/or training is very rare even in children with hearing impairment with good language skills. Using rhythm training strategies in the intervention of children with hearing impairment may prove to improve their temporal predictions, which can facilitate better speech perception and prediction (Hidalgo, Falk &Schon, 2017). This can eventually, result in a more natural speech with increased speech intelligibility.

6. **Conclusion:**

The current study identified the speech rhythm of CWHI to have higher nPVI and rPVI values, which could be another contributing factor to the reduced speech intelligibility. The supra segmental intervention strategies selected could aim at producing a lesser or rather consistent variations in the speech of CWHI, which could improve the naturalness of speech

7. References

[1] Calvert, D. (1962). Deaf voice quality: a preliminary investigation. Volta Review, 64: 402-403

[2] Kent, R. D., & Kim, Y. (2008). Acoustic analysis of speech. *The Handbook of clinical linguistics*, 360-380.

[3] Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850-11857.

[4] Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm:"Syllable-timing"in Singapore English. Language and Speech, 43, 377–401.

[5] Merin, J. (2016). Speech rhythm in Malayalam-speaking children with hearing impairment. (Unpublished master"s dissertation). All India Institute of Speech and Hearing, Mysore, Karnataka.

[6] Osberger, M. J., &McGarr, N. S. (1982). Speech production characteristics of the hearing impaired. In *Speech and Language* (Vol. 8, pp. 221-283). Elsevier.

[7] Parkhurst, B. G., & Levitt, H. (1978). The effect of selected prosodic errors on the intelligibility of deaf speech. *Journal of Communication Disorders*, *11*(2-3), 249-256.

[8] Rosenhouse, J. (1986). Intonation problems of hearing-impaired Hebrew-speaking children. *Language and speech*, *29*(1), 69-92.

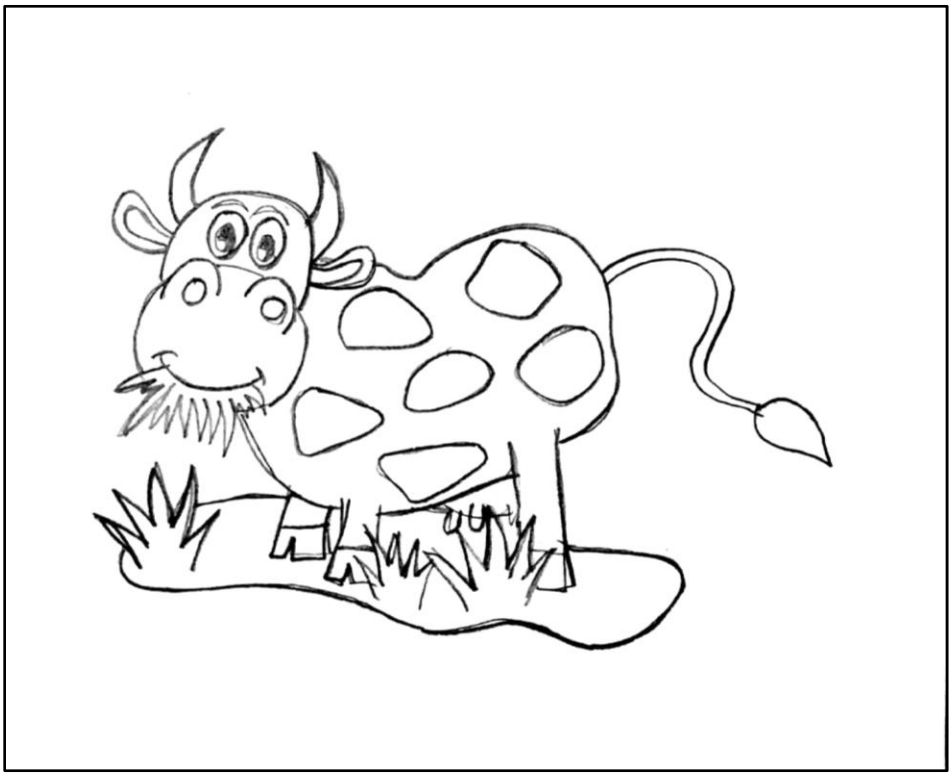
[9] Savithri, S. R., Jayaram, M., Kedarnath, D. &Goswami, S. (2006). Speech rhythm in Indo Aryan and Dravidian languages. *Proceedings of the International Symposium on Frontiers of Research on speech and music*, 31-35.

[10] Savithri, S. R., Johnsirani, R. &Ruchi, A. (2008). Speech Rhythm in Hearing-Impaired Children. *AIISH Research Fund Project*.

[11] Yoshinaga-Itano, C. (2003). From screening to early identification and intervention: Discovering predictors to successful outcomes for children with significant hearing loss. *Journal of deaf studies and deaf education*, *8*(1), 11-30.

Example of a picture representation and the corresponding sentence stimulus.

/paʃU pUlI^ ʈɪn̪n̪U/



Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Meaning Making through 'Music' and 'Emotion' in Bengali Children's Rhymes *(Choras)*

Rajoshree Chatterjee, Jayshree Chakraborty
IIT Kharagpur, India

ARTICLE INFO

Article history:
Received 06/03/2020
Accepted 06/08/2020

Keywords:
Choras,
music,
emotion,
meaning-making, discourse comprehension

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

The study takes up the rare breed of Bengali Elementary Choras because of their quintessential idiosyncrasies, and attempts to explain how 'music' and 'emotion' act as dominant determinants of meaning making. These apparently "incohesive" and "discontinuous" verses are targeted towards little children with underdeveloped semantic, linguistic and/or cognitive ability or schemata. The paper claims that the children appreciate or grasp these elementary Choras by acknowledging and directly experiencing the basic intent of the discourse (which is to provide amusement and positivity). The representatives of such intended emotions of pleasure and elation are the various musical devices embedded in the verses that operate together with the warmth of motherese, bringing about coherence to the listener. The analysis is upheld by different models of Discourse Semantics and Discourse Pragmatics, especially the theories of Teun A. van Dijk and Walter Kintsch which signify the various aspects of discourse and their relevance in the process of comprehension. The study concludes that musical elements in the verses acting as bearers of the aimed emotions of glee and gratification in the children not only contribute to meaning making but also justifies the visible discontinuities.

1. Introduction

The present study attempts to explain how music and emotion serve as primary sources of connectivity in Bengali elementary *Chora*s (children's nursery rhymes) for their target audience, who are little children with little or no schematic competence or world knowledge. The process of comprehension or realizing connectivity in texts has been traditionally understood in terms of coherence and cohesion, where coherence refers to deciphering the intended meaning given out by the overall unity of discourse, while cohesion refers to the connectivity operating within the linguistic composition. The need to focus on the legibility of children's *Chora*s arises because of their very unique and quintessential peculiarities. The primary *Chora*s under the present discussion, brim with associations, depiction and descriptions of one's surroundings and locale which are better experienced than understood, unlike concrete concepts (semantics/ pragmatics) contained in adults' texts, that need to be processed and comprehended. At the level of ideas,

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Rajoshree Chatterjee
Email: rajoshri.chatterji@gmail.com

such basic verses are full of cohesive gaps and discontinuities; nevertheless, they are enjoyed and relished by their target audience as well as adult readers. The reason why they are appreciated by their target audience despite their textual cohesive gaps can be explained if understanding or comprehension in this context is measured in terms of meaning making, i.e making sense of the entire discourse and not just the verses. It is thus witnessed that Elementary *Chora*s do not follow the expected traditional approach of deciphering the textual and contextual meaning; instead, they are built upon 'making sense' of the entire discourse. The intended purpose of the total discourse and/or the communicative intent of the speaker is to make the little children feel amused while interacting with the natural surroundings. The major carrier or device of this proposed positivity and amusement is the 'emotion' induced by the verses through the various textual accessories of music embedded in the verses. The paper therefore throws light on the unique genre of Bengali children's elementary *Chora*s identifying music as the major source of producing emotions of amusement and positivity, which serves as the prevalent factor of connectivity. Before going into the analysis of textual devices of music, providing emotions of positivity and enjoyment, a description of *Chora*s, in general, and elementary *Chora*s, in particular, is in order.

Chora, the Bengali oral folkloric counterpart of nonsensical verses for children, is a type of interactive discourse which transmits major elements of enjoyment ranging from endearment, recreation and amusement to the sense of cultural consciousness and social understanding. Though *Chora*s may be found in the written form today, they have been a notable part of oral literature and passed on for generations, as they retain all the characteristics of oral interactive discourses. They can be defined as short, rhyming, naturalized oral verses meant for children, generously flavored with Bengali cultural connotations. They are uncomplicated, elementary and apparently meaningless rhythmic texts which often seem to be manifestations of the characteristics of child-directed speech. The register of *Chora*s is a simple, informal language of day-to-day interaction, in close personalized settings, generously sprinkled with intermittent cohesive gaps and discontinuities surfacing at every level, as a part of their peculiarity. These gaps may appear 'in the form of unrelated and disjointed ideas, non-words and/or nonsensical phrases, lexical cohesive gaps surfacing through personification of animate or inanimate objects, actions deviant from the normal or accepted course of events and semantically unrelated words grouped together based on their similar rhyme scheme' (Chatterjee & Chakraborty, 2019). Elementary *Chora*s, because of their above mentioned inherent traits, are witnessed to creatively dodge comprehension in their target audience, benefitting from their underdeveloped schematic competence. However, regardless of the cohesive gaps operating in different stages of the genre, they are found to be thoroughly enjoyed and appreciated by the little children despite their lack of prior textual and/or contextual awareness.

The present research proposes how appreciating and grasping 'elementary' *Chora*s targeted towards very little children particularly is very different from their 'advanced' counterparts. While the embedded textual and contextual meaning, the thematic content and/or topic continuity in advanced *Chora*s can be fathomed by children, whose gradually evolved cognitive world aid them to master linguistic and semantic skills, it is not the same for the little children who are yet to sharpen their acumen. For them, the process of grasping and appreciating the text is dependent on their ability to identify only the main motive or purpose of the discourse. The purpose of such discourses is to familiarize the little children with their immediate surroundings in a joyful manner so that they not only associate themselves with their surroundings but also

develop an affinity with them naturally. This process is made possible or felicitated in such discourses because these *Chora*s are purely interactive in nature, where children directly participate in the discourse and endure the effect of interaction naturally and simultaneously. Based on such observations, the study proposes that the term 'comprehension' should be replaced by 'meaning-making' in such interactive discourses. This process of 'meaning-making' comforts the children of the target age group to witness and directly experience the aimed pragmatic intent of the discourse even when they are not capable of decoding the textual and/or contextual meaning. The analysis thrusts on how this unique nature and essence of the discourse triggers us to think and define the process of a text's intelligibility and cognizance erected over musical machineries of enjoyment and amusement.

2. Objective

The objective of the paper is to show that (i) appreciation of the elementary *Chora*s by their target audience is based on their understanding or experiencing the basic intent of the discourse (producing enjoyment or positivity), and (ii) this process of meaning making is activated through various musical devices embedded in the text which serve as the common factor of connectivity by virtue of their being producers of emotions of positivity and pleasure.

3. Theoretical Background

Meaning making is an important feature for a text's acceptability, influence and prominence. Following the traditional views of Haliday and Hasan (1976), 'the concept of cohesion is a semantic one; it refers to relations of meaning that exist within the text and that define it as a text'. The intelligibility of a text was typically thought to rest solely on unity at the textual level, visible through the in-text cohesive links (cohesion) that was considered to contribute to overall coherence. This led to the view that cohesion of a text causes the overall coherence. However, some researchers have also claimed that though cohesion as a textual property of a text is a crucial aspect towards making a discourse meaningful, it is not indispensable. Cognizance is also brought about by unity at the contextual level, where the target audiences' prior schemata helps to connote meaning, making coherence fundamental and dominant. Linguists like Givon (1983) talk about discourse unity in terms of topic continuity and propose a three-level framework for topic continuity: thematic continuity, action continuity, and topics/participants continuity. Similarly, Blakemore(1987, 1992) says that interpretation of a discourse is not always based on the propositions explained but rather on "non linguistic and contextual features" accompanying the discourse. In similar lines, according to Schiffrin (1987), coherence in a discourse is dependent on "a speaker's successful integration of different verbal and nonverbal devices to situate a message in an interpretive frame and a hearer's corresponding synthetic ability to interpret such cues as a totality in order to interpret that message". Such views jointly prompt cohesion to be the effect of coherence as opposed to its counterview. In Bengali elementary *Chora*s too, the cognitive capacity of the target audience limits us to believe that discerning such verses is not possible by acknowledging their textual content. Interpretation of this interactive genre thus pursues the latter approach, where the pragmatic intent of the discourse of producing positivity and amusement in the little children, plays the principal role to provide comprehensive connotation of the verses thereafter ruling out the visible textual discontinuities.

4. Research Methodology

For analyzing the data, the Discourse Semantics methodology which includes knowledge integration and discourse management has been chosen. According to discourse management, the speaker provides and presents information to the listeners depending on their need and cognitive ability. In our analysis, we propose that the speaker first identifies the purpose of the discourse and accordingly selects information such that the listener can make sense of what is communicated. In the context of elementary *Chora*s, considering the listeners cognitive ability, we suggest that the speaker focuses only on those aspects (e.g. music, rhyme, rhythm etc.) which help the listeners understand and experience the 'purpose' or 'rhetorical management' of the discourse. Our analysis in the paper draws its support from various models of Discourse Semantics and Discourse Pragmatics, but the main idea of identifying the purpose of discourse working as the unifying factor in the text is inspired by the works of Teun A. van Dijk and Walter Kintsch, namely Kintsch and Dijk (1978), Kintsch (1988) and van Dijk and Kitsch (1983).

5. Analysis

Connectivity in the majority of texts, specially the texts meant for adults, has been considered to be the property of the text and context both. However, in the present discussion, from a deeper insight into the genre of such discourses and the capacity of the target audience, we strongly believe that meaning in such discontinuous verses, meant for children with little linguistic and semantic skills, is not a textual attribute; properties beyond words such as the language play of music and rhythm intertwined in the verses and visuals arousing excitement and of motherese, work together to bring about emotions of exhilaration and elation. Based on this, we propose that in elementary *Chora*s connectivity is realized when the children identify the common intent of the discourse. Our analysis focuses on those textual and extra textual devices which contribute to this target. This fervor of affection and joy, produced with a bunch of representational overlapping of the textual and contextual features of comprehension in *Chora*s including *sound* and *music, rhyme, rhythm* and *visuals* will be analyzed in the following section.

*5.1 Emotions in Elementary Chora*s

The manifestation of the *Chora*s is through colloquial and homely Bengali wording for the most obvious reason of their target audience, who have little comprehension abilities. The medium is one of feeling and experiencing, as the expressions are emotive in nature and powdered bountifully with cultural jargon, helping the readers to identify and endure the words. *Chora*s are composed in such a way that even if the writer does not explicitly indicate emotions, they are expressed and induced in the reader. Through such affective use of texture, the writer communicates a lot of sentiments and subtly guides the little audiences about the right and wrong doings while keeping the enjoyment factor of the *Chora*s intact. Interjections, endearing terms, words expressing love, warmth and fondness all fall under this unique medium of *Chora*s. Some such emotive words and gestures are found in the lines '*Aay aay chaand mama, tip diye ja..chand-er kopale chand, tip diye ja..', or 'Bhor holo dor kholo …khukumoni otho re…' or 'haathi naachche...Ghora nacchche…Shonamonir bey'.* Children rarely understand the semantic meaning of the words, but they do understand the 'language of love' evident from the tone and the gestures of the recitor. The child can make out if he/she is being scolded or being pampered – which suggest that it is through the medium composed of the non-semantic and nonverbal elements, emotions, gestures, expressions, vocalics, warmth and closeness shared with the narrator, that they construct meaning and enjoy and 'live' the *Chora*s.

In the context of elementary *Chora*s, the participants of the discourse are majorly children and their mother, grandmother or their (female) caretaker. It is essential that the bond between the participants is one of love, warmth, care, tenderness, compassion and affection. The essential purpose of the *Chora*s is that they are often used as lullabies, by the mother or the caretaker to soothe or compose the child, coax it to eat, or relax it to sleep, which portray *Chora*s as manifestations of Motherese. Some examples of affectionate words found in *Chora*s are '*Shonamoni', 'Khoka', 'Shona' 'Khukumoni' Chaand' 'Babu' 'Khuku',* all of which are used in very close and personalized context giving out the endearing and tender tone of the verses, helping the participants (the child and the caregiver) bond emotionally. Such a dear relation is also reflected by personifying inanimate objects like *Chaand Mama, Shujji Mama* etc. Their interpersonal relationship of warmth and affection pave the way for the child to enjoy the verses as well as unconsciously helping them to acquire language. In addition to this, it must not be forgotten that *Chora*s have a very interactive mode of communication, such that the participants involved in the process can directly experience and participate in the medium despite comprehending the direct semantics of language.

The Emotion in the form of shared positivity and pleasure produced in the little children, serves as a binding effect by uniting the entire class of *Chora*s under discussion. The propagators of this pragmatic motive of inducing affection and compassion in the children, is the underlying music embedded in the elementary verses, which by serving as messengers of enjoyment and amusement also rationalize the visible cohesive gaps in the discourse.

Motherese or *Child Directed Speech* - serves as the basic texture of *Chora*s. Motherese is as the medium of Register in *Chora*s is much evident from its textural characteristics and is defined as "a simplified form of language used (especially by mothers) in speaking to babies and young children, characterized by repetition, simple sentence structure, limited vocabulary, onomatopoeia, and expressive intonation" by the Oxford dictionary. Delivered with a different mix of pitch and intonation, these are different from adult speech and usually involve a lot of over-articulation and rhyme. In *Chora*s, it is manifested through gesture, prolonged accentuated tones and intonations, emotive words, affective associations, affectionate kinship terms, various regional/ colloquial/ personalised words, words referring to emotionally loaded social events, personification of inanimate objects etc. Various words like *khoka, khuku, babu, shonamoni, chand mama,* serving as examples to this effect have surfaced throughout our study. Often such words have no meaning, but brim with warmth, affection and positive emotion working as natural stimulants helping children arouse interest. This engrossment helps the child to identify, associate and familiarize with things before they have reached the stage of comprehension. The younger the child, the more exaggerated is the baby talk, which helps to hold the attention of the infant over normal speech, signifying the child's active engagement and attraction to the people engaged in such speech. Motherese therefore serves as an important aspect in making the target audiences of the *Chora*s appreciate its texture, as well as bonds with the participants emotionally.

*5.2 Analyzing Musical Elements in Chora*s

The emotions induced by the *Chora*s rest on the giant shoulders of music, rhythm and their correlating determinants which unite the different unrelated words in a common framework, which we claim, makes perception of the incohesive verses permeable. It is here where the functioning of the idea of 'Language Play' in children's texts comes into limelight, clearing the route and accounting for the operation of music and rhyme in *Chora*s.

5.2.1 Language Play

Language Play according to Neaum (2012) means "playing around with language as in rhymes, jokes, sounds and wordplay, imitation using different voices and all other playful ways of using language". Neaum says that children seem to be drawn to sound and word play, imitation, rhyme and absurdity in language. 'Language play' in the context of *Chora*s are the underlying musical elements - rhyme, rhythm and the play of sounds that add meaning to the otherwise nonsensical words. It is intertwined in the *Chora*s and is one of their fundamental and defining characteristic properties. Children derive delight from the strings of words in the verses, majorly on the basis of their rhyme scheme, rhythm and sound patterns which also aids to boost their language familiarity, without making any denotative or connotative sense. For example: "*Hattimatimtim, taramaathe pare deem, tader kha(n)ra duto sheeng, tara hatttimatimtim".* The absolutely absurd wording of the verse cited above, is majorly built up by the play of sounds and musical elements, which despite any semantic meaning is unconditionally enjoyed by the little children. The words and phrases of the given *Chora* are seen to provide music to the discourse through its patterning sounds and rhyme scheme. The interactive nature of the elementary *Chora*s, aided by the visuals, usher these "nonsensical" expressions to be directly enacted. The meaning thus communicated is constructed in the context and inferred through fond tickling experiences.

5.2.2 Sounds in *Chora*s

The channels of any kind of communication revolve around the human senses of sound, sight and touch. Of these, the auditory vocal mode of communication - sound is expressed through the careful combination of language, style and intonation. *Chora*s are few of the first things that children in a Bengali household are exposed to. The texture of the *Chora*s verbalizes the everyday occurrences of daily life, playing an important role in a child's perception of the surroundings. Being regularly read to, *Chora*s help children to internalize certain sounds (onomatopoeic words like *bhow-wow, bok-bokom*) and later, on confronting a comparably related sound and visual reality, they associate and connect the sounds and the mental images to their actual entities. For example when a child comes across an actual barking dog in real life, it can actually associate with the auditory and mental images of the same that it has acquired from the *Chora*s, making this unique socio-cultural register of these verses highly pedagogically relevant.

5.2.3 Rhyme and Rhythm

A common characteristic that unites *Chora*s with their equivalent counterparts is the use of rhyme and rhythm, which are manifestations of the linguistic notion of phonological parallelism. While Rhythm is defined as the "the systematic arrangement of musical sounds, principally according to duration and periodical stress", Rhyme is "correspondence of sound between words or the endings of words, especially when these are used at the ends of lines of poetry". It is a combination of these two that give the *Chora*s a sing-song flow and a melody which often helps the mother using it as lullabies to compose her child to sleep. Words like *mou, bou; aata, tota, duluni, chiruni,* which surface at the end of the lines are grouped together based on their rhyme scheme despite their un-relatedness. The rhyming words in the *Chora*s forming an integral part of its texture, serve as a cornerstone for the music that flows through the lines which serve as links to relate two words with no semantic connection therefore functioning as a cohesive tie.

6. Discussion

From the above discussion endorsed by the adept study of the genre of *Choras* and the cognitive competence of their target audience, it is supposed that understanding or making meaning in such disjointed verses, is not a textual peculiarity. Coherence instead, is achieved by realizing the communicative intent of the *Chora*s - which is to induce emotions of positivity and pleasure in the children. The extra textual properties such as music and rhythm intertwined in the verses, visuals arousing interpretation and clarity along with the subtle play of motherese, operate together for meaning making. It is perceived that the 'direct experiencing' that takes place through the simultaneous exposure to linguistic and socio-cultural events operating together with expressive language, enhances the ardor induced in the little children. Therefore, the bunch of overlapping attributes of comprehension in *Chora*s include affirmative emotions urged by the play of sounds and music in the verses that helps them to be experienced and enjoyed rather than processed. The sync in the pattern of the music therefore unifies the *Chora*s' spontaneous reflection of the cultural elements and socio-contextual realities that surfaces as apparent textual cohesive gaps in the verses.

7. Conclusion

To sum up, the paper problematizes the question of comprehending apparently "nonsensical" elementary Bengali *Chora*s by little children, and thereafter proposes that meaning making may not necessarily follow the traditional (textual) cohesion leading (overall) coherence approach. Meaning making in this unique genre banks on the exclusive character of such discourses as well as the capacity of their target audience who are yet to sharpen their competence. Intelligence of such verses is therefore through recognizing the pragmatic intent of the genre which is to provide pleasure and amusement to the children. The envoys of the desired enjoyment is the ingrained emotion surfacing through music in the verses, that not only cultivate overall coherence and comprehension but also accounts for the cohesive cracks

References

Blum-Kulka, S., & Hamo, M. (2011). Discourse Pragmatics. In T. Van Dijk, *Discourse Studies: A Multidisciplinary Introduction*. SAGE Publications Ltd.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. British Journal Of Developmental Psychology, 21(3), 335-351. doi: 10.1348/026151003322277739

Carrell, P. (1982). Cohesion Is Not Coherence. TESOL Quarterly, 16(4), 479. doi:10.2307/3586466 .

Chatterjee, R., & Chakraborty, J. (2019). Analyzing Discourse Coherence in Bengali Elementary *Chora*s (Children's Nursery Rhymes). Rupkatha Journal On Interdisciplinary Studies In Humanities, 11(3). https://doi.org/10.21659/rupkatha.v11n3.06

Givón, T. "Coherence in Text, Coherence in Mind." *Pragmatics and Cognition* 1.2 (1993): 171-227. Web.

Gonzalez, A. L. (2016). Music and Language Development: Traits of Nursery Rhymes and Their Impact on Children's Language Development. Diss. California Polytechnic State University.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. Longman.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182. https://doi.org/10.1037/0033-295x.95.2.163

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. https://doi.org/10.1037/0033-295x.85.5.363

Liu, C., & Huang, N. Analysis of the Characteristics of Children's Poetry – Take Shel Silverstein's Poems for Example. National Hsinchu Girls' Senior High School.

Martin, J. (2015). Cohesion and Texture. In D. Tannen, H. Hamilton & D. Schiffrin, *The Handbook of Discourse Analysis* (2nd ed., pp. 61 - 81). Wiley Blackwell.

Neaum, S. (2012). *Language and literacy for the early years*. Sage Publications.

Shapiro, L., & Hudson, J. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. Developmental Psychology, 27(6), 960-974. doi: 10.1037//0012-1649.27.6.960

Tangkiengsirisin, S. Cohesion and Coherence In Text. Language Institute, Thammasat University.

Tartilas, L. (2010). Cohesive Ties in Different Registers (MA). Vilinus Pedagogical University.

Van Dijk, T. (1977). *Text and Context - Explorations in the Semantics and Pragmatics of Discourse* (pp. 93 - 114, 167 - 228). New York: Longman.

Van Dijk, T. A. (1980). The semantics and pragmatics of functional coherence in discourse. *Speech act theory: Ten years later*, 49-65.

Van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York. Academic Press.

Wang, Y., & Guo, M. (2014). A Short Analysis of Discourse Coherence. Journal Of Language Teaching And Research, 5(2), 460 - 465. doi: 10.4304/jltr.5.2.460-465

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

How does musical notes correlate with human emotion? A psycho-acoustic exploration with Indian Classical Music

Medha Basu[a], *Archi Banerjee*[a,b], *Shankha Sanyal*[a] *& Dipak Ghosh*[a]
[a]Jadavpur University, India; [b]IIT, Kharagpur India

ARTICLE INFO

Article history:
Received 23/06/2020
Accepted 06/08/2020

Keywords:
Indian Classical Music,
emotion,
notes,
EEG,
MFDFA

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

The most interesting feature of Indian Classical Music is the existence of *Ragas*. Each *Raga* has its own peculiar ascending and descending movement called the *Arohana* and *Avarohana*. Even if two (or more) *Ragas* are made up of the same notes, the combinational varieties of notes evoke different emotions. In this work, we envisage to study how emotion perception in listeners' changes when there is an alteration of merely a single note in a pentatonic *Raga* and also when a particular note(s) is replaced by its flat/sharp counterpart. Approximately 60 sec recordings were done for two pair of *Raga*s which were chosen in a manner such that they are having difference in only one note keeping all others same. The fractal dimension of the auditory waveform provides a robust nonlinear quantitative parameter with which the two pair of audio clips can be compared. Also, the emotional appraisal from these two pairs were assessed on the basis of psychological listening tests as well as from cognitive response in the form of EEG experiments done on 5 participants. Interesting new results are obtained on how a trivial change in the note structure of a particular Raaga influences human emotion to a large extent

1. INTRODUCTION

Hindustani Classical Music

The Indian classical music (ICM) system is based on the note system which consists of 12 notes, each having a definite frequency. The main feature of this music form is the existence of *'Ragas'*. Each *Raga* is unique, having a definite combination of these 12 notes, and captures a particular mood or emotion. Also, the rhythm of each *Raga* is very specific and unique in every aspect. The presence of 12 notes is not essential in each of the *Raga*; some can have only 5 notes which are

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Medha Basu
Email: medhabasu1996@gmail.com

usually called 'pentatonic *Ragas* or scales'. The *Ragas* which have all the 12 notes are much elaborate in nature and rendition. All the major traditional *Ragas* fall under one of the 10 basic '*Thaats*' which are essentially a particular musical scale or framework. Each *Raga* has a definite musical form or image called the **Swaroopam (**Shetty and Achary, 2009**),** which mainly arises from the notes used, the movements like '*gamak*', '*andolon*' assigned to these notes and the specific patters and progressions in which these notes are presented.

Each *Raga* evokes a particular emotion which can be joy, sadness, anger, disgust, fear etc. if presented in the correct '*laya'* or rhythm. The definition of '*rasa'* is different from being a single emotional state, but a superposition of emotional states. The conceptualization of emotions based on '*navarasa'* is an emotional classification given by Bharata, that suits behavioral studies with Indian arts. According to this classification, the principle *Rasa* or emotions are **Shringar**- love/erotic, **Hasya**- joy/laughter, **Vira**- heroism, **Raudra**- anger, **Bibhitsa**- disgust, **Adbhuta**- wonder and **Bhayanaka**- fear.

The recognition of human emotion can now be studied computationally using different features of Brain Computer interaction (BCI). The musical input provided can be analyzed from various acoustic parameters such as pitch profile, amplitude variations and other quantitative parameters such as spectral energy, spectral skewness etc. Different *Ragas* have distinct values of these quantitative parameters and can be further studied in details from the acoustic analysis. Nonlinear chaos based methods such as fractal analysis proves to be an important tool with which the inherent complexities of the different *Raga* clips can be studied quantitatively. The same technique can be applied to assess the nonlinear non-stationary EEG signals to assess the brain response corresponding to the change in emotional state.

The Electroencephalogram (EEG) is basically the averaged action potential of all the neuronal activities going on in the different lobes of the human brain. The musical application of Brain Computer Interface (BCI) can represent the connections between mental states and music (Miranda and Brouse, 2005; Miranda, 2010; Wu et al., 2010), and detect users' current affective states significantly (Daly et al., 2016). To express the activities of different brain regions, several instruments were used to represent different brain regions and that just make the brain like an orchestra (Hinterberger and Baier, 2005); the voice or music for the left and right channels were deduced by the activities of the respective spheres (Wu et al., 2014). Deriving a quartet from multi-channel EEGs with artistic beat and tonality filtering, we can harmonically distinguish the different states of the brain activities (Wu et al., 2013). The combination of EEG and fMRI provided more information of the brain which can be heard (Lu et al., 2012). To assess the complex non-linear, non stationary EEG signals, a robust nonlinear technique called Multifractal Detrended Fluctuation Analysis (MFDFA) (Kantelhardt, et.al, 2002) have been used which takes into account the widely varying fluctuations present in the EEG signals.

The beauty of Indian music as illustrated above lies in the fact that a mere change in the single frequency of a *Raga* clip changes it to another *Raga* and also the associated emotion changes along

with it. It would be interesting to study the changes in acoustical, perceptual and neural features associated with the change of a single note. In this work, for the first time, we envisage to study how the emotion perception in listeners' changes when there is an alteration of a single note in a pentatonic *Raga* and also when a particular note(s) is replaced by its flat/sharp counterpart. Robust nonlinear methods such as DFA and MFDFA have been utilized to quantify the acoustical signal as well as the brain arousal response corresponding to the two pair of *Ragas* taken for our study. We aim to tackle the problem both from psycho-acoustical and neurological perspective.

2. EXPERIMENTAL DETAILS:

1. A pair of *Ragas* which are having same notes but changing two notes to their respective sharp (or flat) counterparts converts one to another.

   Durga- sa re ma pa dha sa
   Gunkali- sa RE ma pa DHA sa

   In this case the notes '*re'* and '*dha'* of *Durga* is changed to their respective sharp/flat counterparts which change *Raga Durga* to *Raga Gunkali.*

2. Another pair of *Ragas* are chosen such that they are having difference in only one note keeping all others same.
   Durga- sa re ma pa dha sa

   Bhupali- sa re ga pa dha sa

   Here, the note '*ma'* in *Raga Durga*, when changed to '*ga'*, makes the *Raga Bhupali*

For each *Raga*, sound clips of around 50-55 secs were recorded.

All the recorded clips were subjected to human response analysis, for standardization of their emotional context, and then acoustical and bio-sensor analysis were performed on them.**Human Response analysis-**

A human response study was done with 50 subjects of varied age, profession and gender. They were provided with an emotion chart of the basic 4 emotions- joy, sorrow, anxiety and calmness, and were asked to mark the clips with their perceived emotional arousal.

Acoustical Analysis- Initially, the four music clips were subjected to linear acoustic analysis in the form of pitch profile analysis and fundamental frequency assessment, while also looking at the correlation between phrases from different pair of clips.

The following figures show how the fundamental frequency corresponding to the different notes change which manifests in the change of *Raga* and hence the emotional appraisal changes.

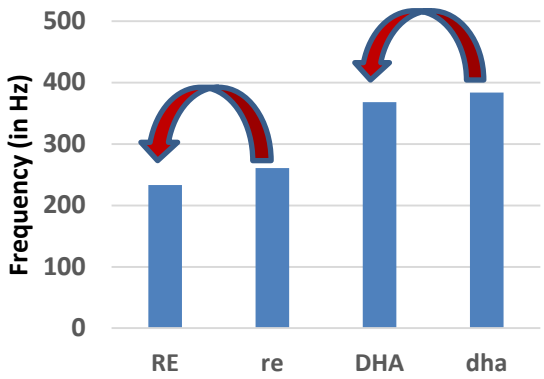


Fig. 1: The difference between musical scale of *Durga* and *Gunkali*

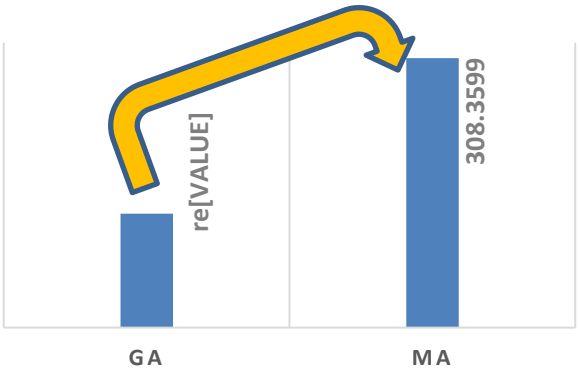


Fig. 2: The difference between musical scale of *Durga* and *Bhupali*

Moreover, the acoustical signals being non-stationary and nonlinear and nature demands assessment with the help of robust methods which takes into account these properties of the signals. Hence complexity values corresponding to each clip was evaluated and these values were compared with their counterparts in bio-signal analysis.

EEG Analysis- 5 (Five) subjects participated in an EEG study which involved the target clips as well as four other clips (i.e. a total of 8 clips) which have been used as a distractor so that there is no biasing on the obtained EEG data. The obtained EEG data have been analyzed with the help of robust nonlinear techniques from which the complexity values associated with each clip have been evaluated and this has been used as a parameter with which the brain correlates corresponding to each clip has been characterized.

3. EXPERIMENTAL METHODOLOGY

1. Emotional Response Study

50 participants (M=23, F= 27; mean age= 25 years, SD=2.5 years) were asked to rate the 8 clips played randomly in a scale of 1 to 10, as per the emotional chart given below:

Raga	joy	anxiety	sorrow	Calm
Clip 1				
Clip 2				
Clip 3				
Clip 4				
Clip 5				
Clip 6				
Clip 7				

Multiple marking was also allowed, taking into consideration the ambiguous nature of Indian Music. Radar maps were drawn from those data for a better study.

2. Pitch listing, correlation, Hurst values computed.

Audio clips of about 55-60 seconds were recorded for the given sets of *Ragas*. As seen earlier, from the musical scales of the two *Ragas*, for i) the notes re and dha in *Durga* have been converted to their sharp counterparts in *Gunkali*, while for ii) the two scales differ only in one note i.e ma in Durga which becomes ga in Bhupali.

The recordings were done in such a way that the whole musical scale was covered, including the important phrases i.e. the combination of notes unique for that particular *Raga*. Now, from both the clips, such zones were selected which contained almost the same notes. Of course, in this case, phrases which had re or ma in Durga had the sharp counterparts in Gunkali while for ii) the phrases differed only in one note i.e. ma and ga for Durga and Bhupali respectively. But apart from this, no other difference was allowed in choosing the zones. The length of these selected zones ranged from 3-5 seconds.

Now, such pairs of clips which had the same notes were analysed using Praat software. First, the pitch listing was done for such phrases and then the intensity listing. From the pitch listing, the pitch correlation was calculated using Pearson Correlation. From pitch listing, the graphs were drawn and compared for clips having same notes but differing only in the sharp or flat counterparts.

3. EEG analysis.

5 naive listeners (3M, 2 F, average age = 22 yrs SD=1.3) were chosen for the EEG analysis. The main objective of this study is to observe how the brain responds to a subtle change of a note in the music clip, which is nothing but a mere change in the frequency. Hence, the frontal, occipital, parietal and temporal lobes were selected for the study. EEG was done to record the brain-electrical response of two male subjects. Each subject was prepared with an EEG recording cap with 19 electrodes (Ag/AgCl sintered ring electrodes) placed in the international 10/20 system. Figure 3 depicts the positions of the electrodes. Impedances were checked below 50 kΩ. The EEG recording system (Recorders and Medicare Systems) was operated at 256 samples/s recording on customized software of RMS. The data was band-pass-filtered between 0.5 and 35Hz to remove DC drifts and suppress the 50Hz power line interference.

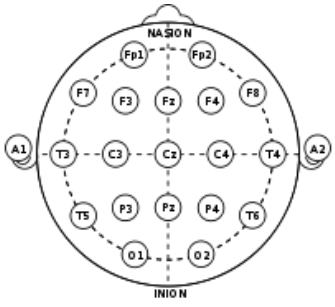


Fig. 3: Position of electrodes

A template of audio clips of about 12 min duration was made in the following manner-

Silence- 1 min => Clip 1- *Raga Durga* (50 sec) => Silence- 30 sec => Clip 2- Distractor 1 *Raga Hameer* (50 sec) => Silence- 30 sec => Clip 3- Distractor 2 *Raga Jog* (53 sec) => Silence- 30 sec => Clip 4 - *Raga Gunkali* (54 sec) => Silence- 30 sec => Clip 5 - Distractor 3 *Raga Marwa* (57 sec) => Silence- 30 sec => Clip 6- *Raga Bhupali (*54 sec*)* => Silence- 30 sec => Clip 7- Distractor 4 *Raga Kedar* (52 sec) => Silence- 1 min

EEG datas were collected for 5 participants and studied in detail using the robust MFDFA technique. From, each and every clip, the exact phrases were extracted which contained the note variation, which ultimately reflected in the output emotional change and the multifractal complexity were computed for the following electrodes: **FZ, FP1, FP2, F3, F4, T3, T4, O1 and O2** For such sets of sections for both the *Ragas*, multifractal width were calculated. This whole process was carried out for all the main phrase-pairs of the *Raga* sets for all the electrodes mentioned above.
The Multifractal Detrended Fluctuation Analysis (MFDFA) was performed on the selected phrases as per the algorithm proposed by Kantelhardt et.al.(2002) and according to the methodology in Sanyal et.al (2016). The method gives output in the form of "w" or the multifractal spectral width, which is essentially the complexity associated with the acoustic or EEG signal.

4. RESULTS AND DISCUSSIONS

1. Human Response Analysis

The following table gives the human response data obtained from an experiment conducted on 50 participants, who were made to listen to the 7 clips in exactly the same order as the EEG experiment was conducted:

Table 1: Emotional response obtained from 50 participants

RAGA	JOY	ANXIETY	SORROW	CALM
Durga	24	1	7	21
Hameer	31	6	8	9
Jog	9	9	14	22
Gunkali	2	16	30	6
Marwa	9	16	18	11
Bhupali	3	5	16	27
Kedar	18	9	12	12

From the table, we plot the emotional response corresponding to the two target pairs in the form of radar graphs as shown below:

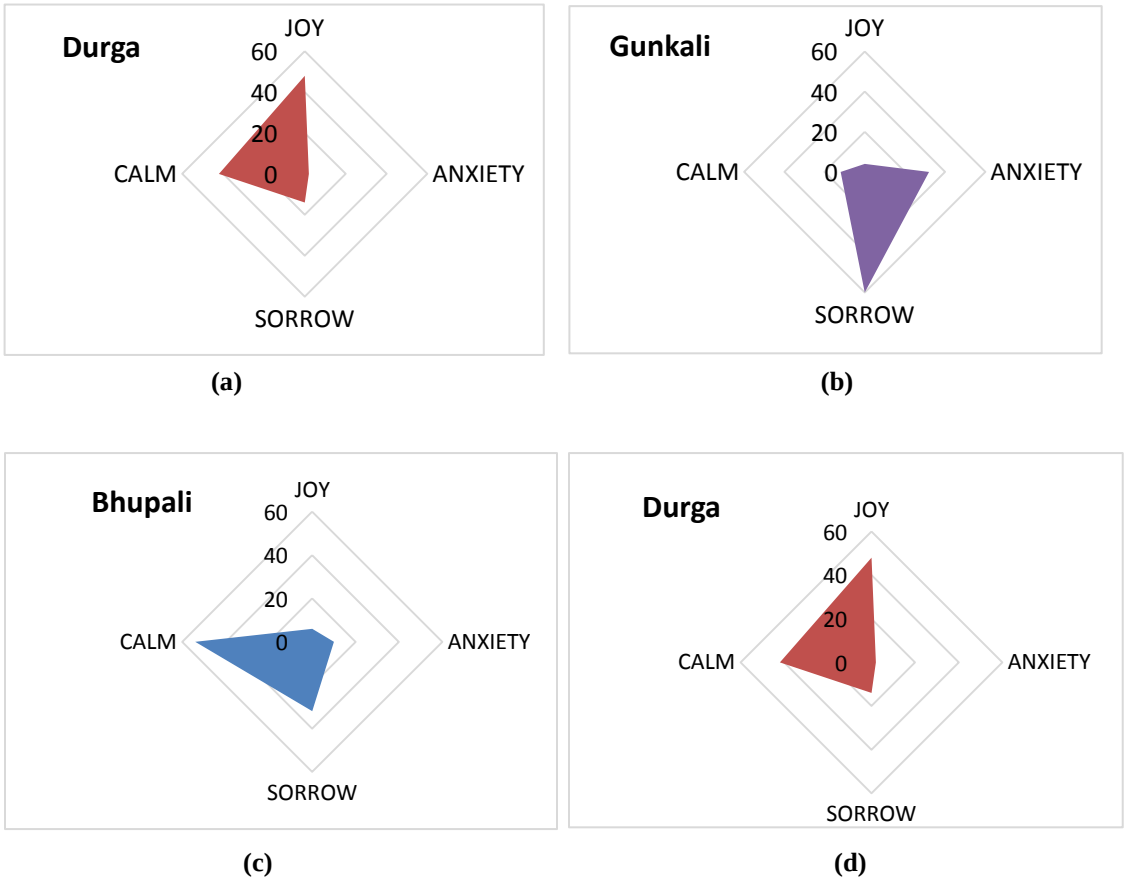


Fig. 4 (a-d): Radar plots for the emotional response of the two pair of chosen *Ragas*

It is clear from the plots that the change of a single note manifests in a complete change in emotional appraisal at the perceptual level of the listeners. How do these changes reflect in the acoustical and neural level would be our aim of study in the next few sections.

2. Acoustical Analysis

From the selected phrases in the pair of *Ragas*, the pitch values were computed. Pearson correlations were calculated from these pitch lists for the phrase pairs and plotted.

CASE 1 - Durga and Gunkali

Phrase (Durga - Gunkali)	Correlation Coefficient
re - RE (P1)	-0.74
dha – DHA (P2)	-0.11
sa - sa (P3)	0.86
ma re ma – ma RE ma (P4)	-0.06
ma pa – ma pa (P5)	-0.02
ma pa dha pa ma – ma pa DHA pa ma (P6)	0.50
ma pa dha sa – ma pa DHA sa (P7)	0.86
sa dha ma pa – sa DHA ma pa (P8)	0.76

ma re – ma RE (P9)	-0.001
dha sa – DHA sa (P10)	0.47

Table 2: Correlation Coefficient for Durga-Gunkali

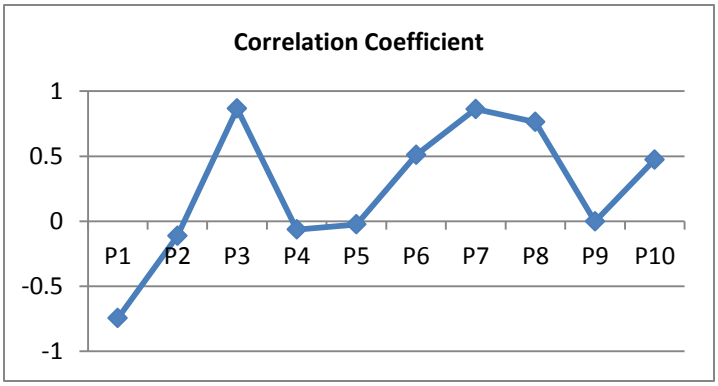


Fig. 5: From the pair it is seen that the *'Shuddha'* and *'komal' re* are anticorrelated while the *'Shuddha'* and *'komal' dha* are loosely correlated but the degree of anti-correlation is much higher in *"re*". The tonic "*sa*" manifests the highest degree of correlation. Any phrase having the *re'*s thus have higher values of anticorrelation than phrases having the *dha'*s.

CASE 2 - Durga and Bhupali

Phrase (Durga - Bhupali)	Correlation Coefficient
ma re dha sa – ga re dha sa (P1)	0.89
ma sa re pa – ga pa (P2)	0.043
ma pa dha sa – ga pa dha sa (P3)	0.41
sa dha ma re pa – sa dha pa ga pa (P4)	0.28
dha pa ma pa – pa dha pa re ga (P5)	0.45
dha pa ma re – pa ga re (P6)	0.20

Table 3: Correlation Coefficient for Durga-Bhupali

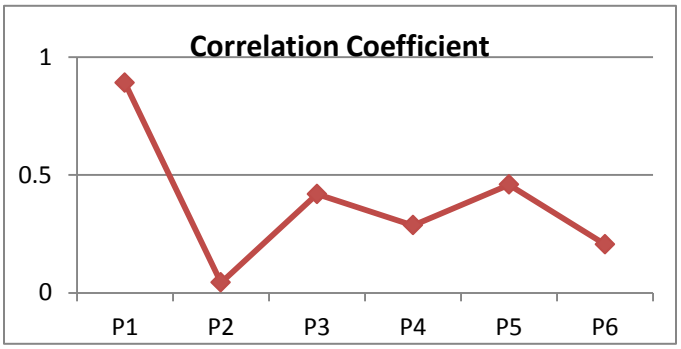


Fig. 6: For this pair, no '*komal*' or sharp notes have been used for any of the *Ragas* and hence there is no observable anti correlation amongst any pairs.

The degree of correlation varies from low to high, low correlation is observed for pair 2 and 4 whenever there is a change with the "*re*" note, while a higher degree of correlation is observed in pair 1 and 3, whenever there is a change from "*ga*" to "*ma*". Thus the correlation between pitch profiles of different pair of acoustical segments can prove to be an important cue while assessing the emotional variation at the source level.

Next, the multifractal width was computed and plotted for these phrases as described above:.

CASE 1- Durga and Gunkali

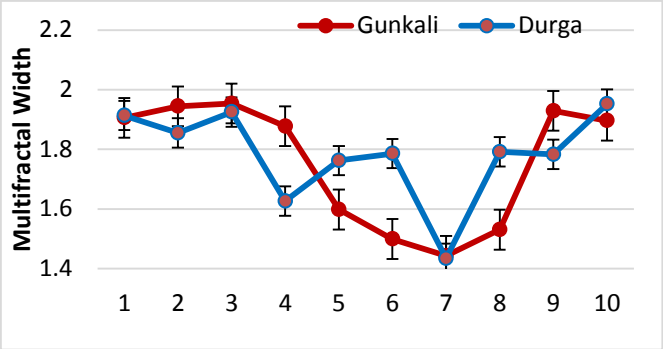


Fig. 7: From the figure, it is seen that the complexity pattern of the single note frequencies are almost similar. The complexity pattern becomes different as the note structure varies for the two *Ragas*. Thus it can be empirically said that it is the subtle variation in the longer note structures that catalyzes the change in human emotion, if it is assumed that the complexity pattern plays a major role in catalyzing human emotion.

CASE 2- Durga and Bhupali

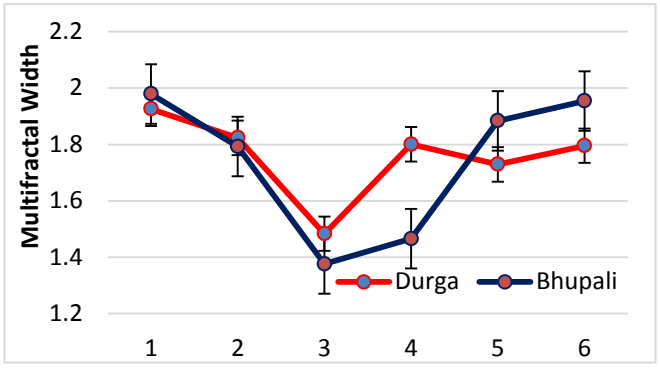


Fig. 8: In this case, again we find that the scaling pattern is almost similar in most of the cases. This can be attributed to the fact that almost similar note patterns are used in most cases, with a change in only a single note.

Again in case of longer note combinations, we find that the scaling pattern in the two different *Ragas* change. An interesting observation is that even though the scaling pattern changes to a small extent in the two pairs, emotional appraisal changes to a large extent, which is evident from the human response data collected earlier. This shows that how, even a small change in the physical properties of the source creates a large change in the output.

3. EEG Analysis:

The multifractal width was calculated corresponding to each of the electrodes when the respondents listened to the chosen phrases in each of the two pair of *Ragas*. The following plots give the brain response to the chosen pair of two *Ragas*. P1... P5 (as in **Table 4**) indicates the multifractal width corresponding to the different phrases as mentioned in the earlier sections.

Table 4: Phrase combinations taken for the two *Ragas* in EEG analysis

PHRASE	NOTE-COMBINATION IN EEG Gunkali- Durga
P1	RE DHA sa – re dha sa
P2	ma RE ma pa – ma re ma pa
P3	ma pa DHA sa- ma pa dha sa
P4	sa DHA ma pa- sa dha ma pa
P5	ma RE sa DHA sa- ma re sa dha sa

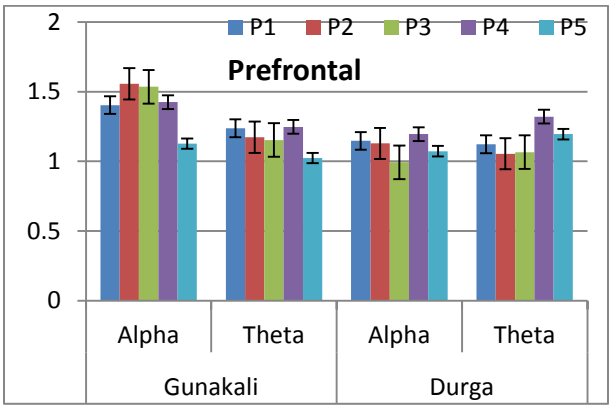


Fig. 9

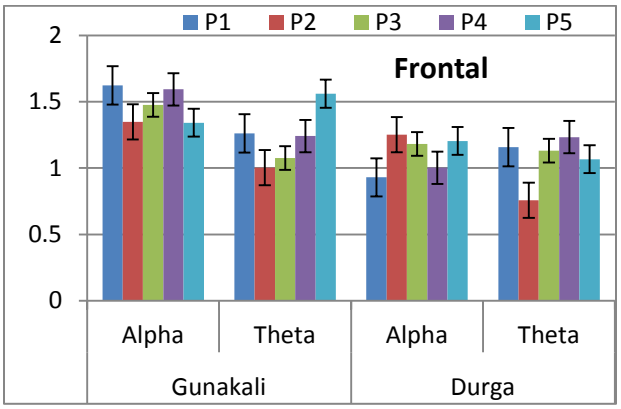


Fig. 10

In both frontal and prefrontal lobes, it is seen (from **Figs. 9 and 10**) in general that the EEG signal complexity for both alpha and theta frequency region decreases when the emotional appraisal changes from negative to positive, i.e. the phrases in *Raga Gunkali* changes to phrases in *Raga Durga.* In the frontal region, changes for P1 and P4 is quite significant in alpha domain, while for theta domain change in P5 is most significant. In case of prefrontal lobe, however the change in P2 and P3 alpha is most significant.

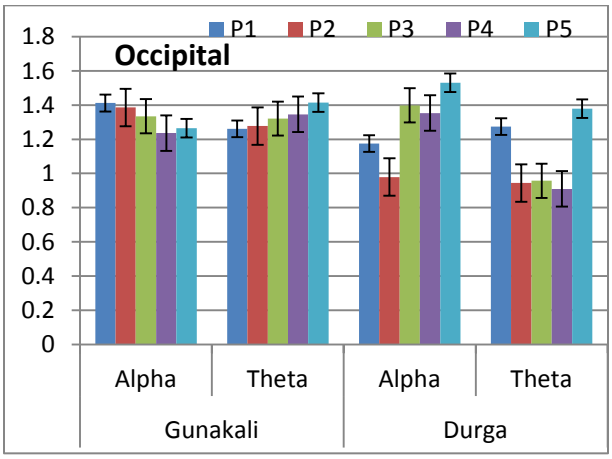


Fig. 11

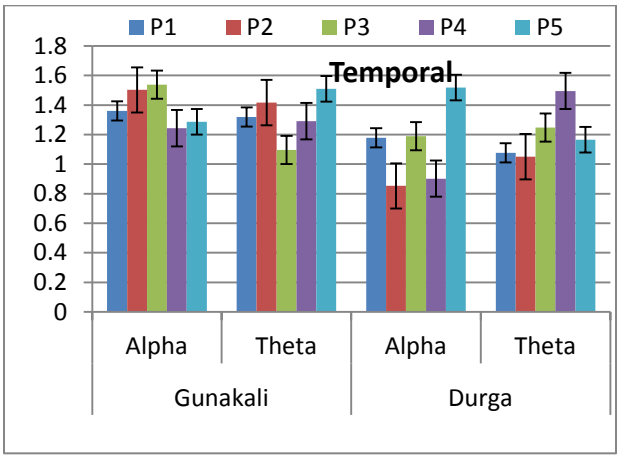


Fig. 12

Figs. 11 and 12 reveal that in both occipital and temporal lobe, the response is somewhat different from what we have seen in the frontal lobe. The alpha frequency region registers a significant decrease in P1 and P2 while for the other phrases, complexity increase in the occipital lobe. For the theta frequency range, however there is a considerable decrease in complexity for all the phrases. In the temporal lobe for phrases 2 and 4, there's a sharp decrease in alpha complexity, when the emotional appraisal changes from sorrow to joy. The following figures show these changes for another pair of *Ragas*, in which there is change of a single note from lower to higher octave (**Table 5**).

Table 5: Phrase combinations taken for the two *Ragas* in EEG analysis

PHRASES	NOTE COMBINATION IN EEG Bhupali – Durga
P1	ga re dha sa – ma re dha sa
P2	ga pa – ma sa re pa
P3	ga pa dha sa – ma pa dha sa
P4	sa dha pa ga pa – sa dha ma re pa
P5	pa ga re – dha pa ma re

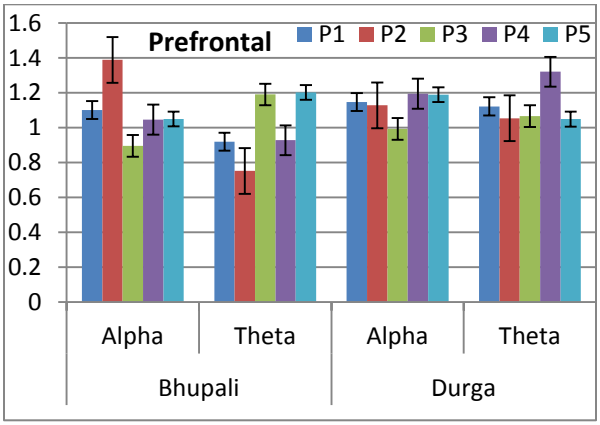


Fig. 13

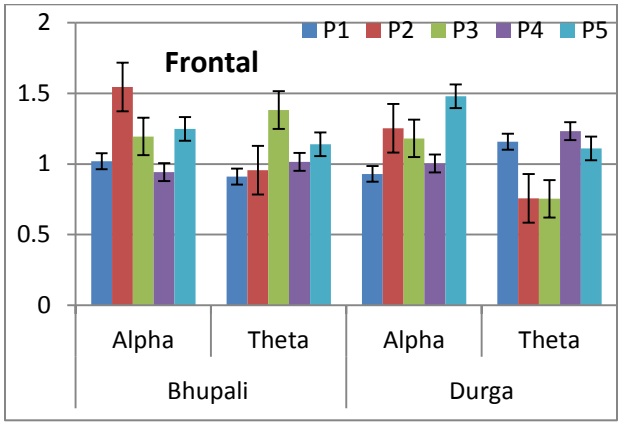


Fig. 14

It is evident from the figures (**Fig. 13 and 14**) that in case of *Raga* Bhupali and *Raga Durga*, the change in brain signal complexity is not that pronounced as in the case of previous pair. This can be attributed to the fact that the contrast in emotional appraisal in this pair is not that significant as compared to the other two. In this case, however we see much more phrase specific arousal in both the alpha and theta region of the pre-frontal and frontal lobe. While P2 records a decrease in alpha complexity of both frontal and prefrontal region, P4 records an increase in theta complexity in both these regions.

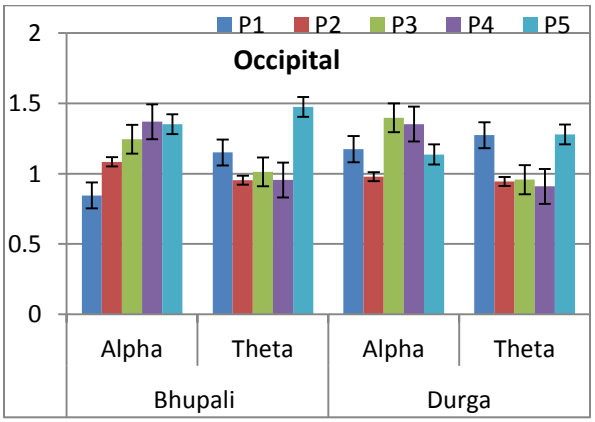


Fig. 15

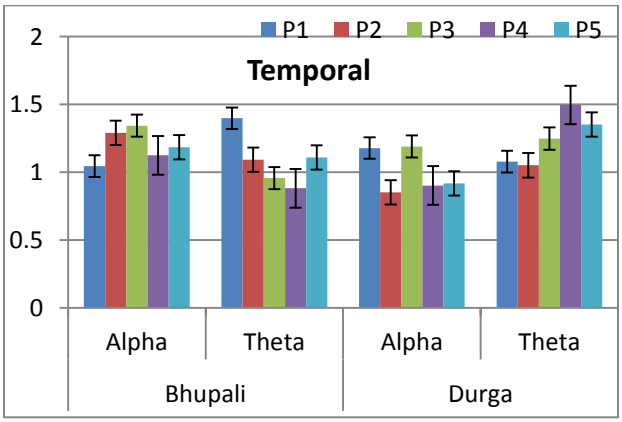


Fig. 16

For the occipital and temporal lobe, again we see that the response is very selective on the different phrases concerned. While for most phrases, both alpha and theta signal complexity increases when the musical emotion changes in the occipital lobe, in the temporal lobe alpha signal complexity decreases in general while theta complexity increases.

In this way, we have developed a method, which can monitor the brain arousal response at such a precise level that when the emotional context of a musical piece changes due to a single note, the same is reflected in the EEG response. The brain response is more precise and distinguishable when the emotional appraisal is strongly different in the musical pieces, while it is varied when the two pieces are close in terms of emotion arousal.

5. CONCLUSION

In this study, for the first time we try to correlate the brain response when there a change in single note cause a change in the *Raga* as well as a change in the emotional context of the *Raga.* Robust linear and nonlinear techniques have been used to quantify the perceptual, acoustic and neural correlates obtained from the study.

The study presents the following interesting conclusions:

1. By making minor changes in the note structure of a *Raga*, a different *Raga* is created i.e. using variation of merely a single note, emotional appraisal is totally changed.
2. This difference in emotional response is more pronounced in case of sharp notes transitions than in '*Shuddha'* note transitions.
3. Using the pitch profile analysis, we found that the *'Shuddha'* and *'komal'* notes are anti-correlated.
4. The multifractal exponents are found to be almost similar for the two pairs chosen, except for the longer note combinations used, indicating similar long range correlations present in the pairs.

5. The complexity parameters obtained from the EEG analysis show strong response in case of Case 1 when there is complete contrast in the emotional appraisal. The alpha power and theta power decreases considerably in the frontal and prefrontal lobes, while in temporal lobes, phrase specific arousal is seen. In Case 2, however such distinction is absent in all the lobes, the arousal is very much specific to the phrases. This can be attributed to the fact that the human response data showed the emotional arousal in Case 2 is not strongly opposite to each other.

In conclusion, it can be said that the study for the first time presents a scientific analysis on how the acoustic, perceptual and neural features change when the emotional appraisal is changed due to the change of a single frequency in a particular *Raga,* Analysis with greater number of samples will lead to more precise information in this domain.

6. ACKNOWLEDGEMENTS:

One of the authors, acknowledges the Department of Science and Technology (DST), CSRI, Govt. of India for providing (CSRI-PDF/34/2018) the CSRI Post-Doctoral Fellowship to pursue this research work. Another author, acknowledges the JU RUSA 2.0 Post-Doctoral Scholarship Scheme for providing the Post-Doctoral Fellowship to pursue this research (R-11/557/19).

REFERENCES:

Daly, I., Williams, D., Kirke, A., Weaver, J., Malik, A., Hwang, F., ... & Nasuto, S. J. (2016). Affective brain–computer music interfacing. *Journal of Neural Engineering*, *13*(4), 046022.

Hinterberger, T., & Baier, G. (2005). Parametric orchestral sonification of EEG in real time. *IEEE MultiMedia*, *12*(2), 70-79.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, *316*(1), 87-114.

Lu, J., Wu, D., Yang, H., Luo, C., Li, C., & Yao, D. (2012). Scale-free brain-wave music from simultaneously EEG and fMRI recordings. *PloS one*, *7*(11).

Maity, A. K., Pratihar, R., Mitra, A., Dey, S., Agrawal, V., Sanyal, S., ... & Ghosh, D. (2015). Multifractal Detrended Fluctuation Analysis of alpha and theta EEG rhythms with musical stimuli. *Chaos, Solitons & Fractals*, *81*, 52-67.

Miranda, E. R. (2010). Plymouth brain-computer music interfacing project: from EEG audio mixers to composition informed by cognitive neuroscience. *International Journal of Arts and Technology*, *3*(2-3), 154-176.

Miranda, E. R., & Brouse, A. (2005). Interfacing the brain directly with musical systems: on developing systems for making music with brain signals. *Leonardo*, *38*(4), 331-336.

Sanyal, S., Banerjee, A., Patranabis, A., Banerjee, K., Sengupta, R., & Ghosh, D. (2016). A study on Improvisation in a Musical performance using Multifractal Detrended Cross Correlation Analysis. *Physica A: Statistical Mechanics and its Applications*, *462*, 67-83.

Shetty, S., & Achary, K. K. (2009). Raga mining of Indian music by extracting arohana-avarohana pattern. *International Journal of Recent Trends in Engineering*, *1*(1), 362.

Wu, J., Zhang, J., Ding, X., Li, R., & Zhou, C. (2013). The effects of music on brain functional networks: a network analysis. *Neuroscience*, *250*, 49-59.

Wu, T., Yang, B., & Sun, H. (2010). EEG classification based on artificial neural network in brain computer interface. In *Life system modeling and intelligent computing* (pp. 154-162). Springer, Berlin, Heidelberg.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Finding Geometry in Quantifiers: A Cognitive Perspective

Spandan Chowdhury
Jadavpur University, India

ARTICLE INFO

Article history:
Received 13/04/2020
Accepted 06/08/2020

Keywords:
quantifier,
natural language,
cognitive linguistics,
semantics,
Scope Ambiguity,

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

This paper aims to portray and understand the geometric nature of quantifiers and quantifier scope in natural language. The problem is investigated from the perspective of cognitive linguistics and semantics, along with projective geometry (Veblen and Young, 1918; Hartshorne, 1967) and set-theoretic interpretations. This paper revisits Fauconnier's theory of Mental Spaces (Fauconnier, 1985, 1997) and Gärdenfors' Conceptual Spaces (Gärdenfors, 2000, 2004) and posits a hybrid approach by merging these two theories. Through this hybrid approach, a separate mental space or cognitive construct for the interpretation of quantifiers is logically argued for. However, the space proposed is different, in nature and construct, from the mental spaces which were proposed by Fauconnier. The behaviour of quantifiers is investigated in relation to this new hybrid space. The paper also makes an attempt to explain, with the help of empirical evidence, the phenomena of Scope Ambiguity, Quantifier Raising & relative scopes of quantifiers under the light of this hybrid approach of mental space model of quantifiers using geometry.

1. Introduction

In truth-conditional semantics, the standard way of representing the status of situations is as possible worlds—there is the real world, and there are worlds with situations that are possible but not necessarily actual (Kearns, 2011). Possible worlds are identified with a person's mental attitude (Croft & Cruze, 2004). (Fauconnier, 1985, 1997) proposes an alternative model of representing the status of knowledge in semantic and pragmatic analysis. Fauconnier replaces the notion of a possible world with that of a mental space, and claims that the mental space is a cognitive structure where utterances are interpreted and situations are mapped in the mind of the speaker (and hearer). On the other hand, (Gärdenfors, 2000, 2004) posits the existence of conceptual spaces which give shape to our perception. According to Gärdenfors, our perception

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Spandan Chowdhury
Email: spandan.gyan@gmail.com

has geometrical structure within our cognition and the geometry helps to shape our perception based on quality dimensions.

On another note, in natural language, quantifiers and their scope-taking phenomena have been the center of continuous research. Taking quantifiers under investigation, this paper shall try to investigate whether any geometry can be assigned to the working of quantifiers, and whether scope-taking can be understood as a geometric phenomenon in our cognitive structure. This paper shall also make an attempt to logically establish a theory which merges the approaches of Mental Space Theory (Fauconnier, 1985, 1997) and Conceptual Space Theory (Gärdenfors, 2000, 2004), to posit the existence of a multi-dimensional conceptual space in our cognition triggered by quantifiers in language.

In the following sections, the paper shall introduce the fundamental pre-requisites to build up the hypothesis. Further, it shall briefly revisit mental space theory and conceptual spaces. Following this, the paper shall present the hypothesis and discuss a few supporting examples as empirical evidence.

2. Theoretical Background

This section briefly introduces the theoretical ideas which are a pre-requisite for the subsequent discussion in section 3. The following sub-sections shall introduce the major ideas involved in formulating the hypothesis of this paper. The aim is to discuss the terminologies by referring to the existing literature and gradually unfolding the central idea.

2.1 Concept

Concepts are defined as abstract ideas or general notions that occur in the mind, in speech, or in thought (Carey, 1991). They are the fundamental building blocks of our thoughts and our beliefs. They play an important role in all aspects of cognition. (Margolis & Lawrence, 2012) describe concepts in three different ways:

- Concepts as *mental representations*, where concepts are entities that exist in the mind (mental objects)
- Concepts as *abilities*, where concepts are abilities peculiar to cognitive agents (mental states)
- Concepts as *Fregean senses*, where concepts are abstract objects, as opposed to mental objects and mental states

According to (Murphy, 2002), concepts are mental representations that allow us to draw appropriate inferences about the type of entities we encounter in our everyday lives. Thus, a concept is used by the brain in order to denote a class of things in the actual world. In this paper, the view of concepts as mental representations shall be taken into account. Thus, concepts are the hypothetical internalized cognitive symbols and symbolic structures that represent the external reality.

2.2 Meaning & Conceptualization

Meaning construction is the process whereby language ‘prompts for’ novel cognitive representations of varying degrees of complexity (Evans and Green, 2006). These representations relate to different scenes and aspects of scenes which are conceived, such as states of affairs in the world, emotion and affect, subjective experiences, and so on. Cognitive semanticists treat meaning construction as a process that is fundamentally conceptual in nature. From this perspective, sentences work as ‘partial instructions’ for the construction of complex

but temporary conceptual domains, assembled as a result of ongoing discourse. These domains, which are called *mental spaces*, are linked to one another in various ways, allowing speakers to 'link back' to mental spaces constructed earlier in the ongoing linguistic exchange. From this perspective, meaning is not a property of individual sentences, nor simply a matter of their interpretation relative to the external world. Instead, meaning arises from a dynamic process of meaning construction, which we call *conceptualisation*.

The way we understand language is through the interaction between semantic structures and conceptual structures, which are mediated by various linguistic mechanisms and conceptual mechanisms. Meaning construction processes concerned with the sorts of nonlinguistic mechanisms central to meaning construction, that are fundamentally nonlinguistic in nature, have been referred to as '*backstage cognition*' (Evans and Green, 2006). There are two theories of backstage cognition which are distinct with respect to one another: Mental Spaces Theory and Conceptual Blending Theory. Mental Spaces Theory describes the nature and creation of mental spaces, which are types of small packets of conceptual structure, produced or built up as we think, structure our discourse and talk. On the other hand, Conceptual Blending Theory deals with the integrative mechanisms and networks which are responsible for producing emergent aspects of meaning, that is, meaning that is in some sense novel, and these mechanisms operate over the collections of different mental spaces.

2.3Mental Spaces

Fauconnier defines mental spaces as 'partial structures that proliferate when we think and talk, allowing a fine-grained partitioning of our discourse and knowledge structures' (Fauconnier, 1997: p.11). Mental spaces consist of elements and they are structured with the help of frames and cognitive models. They are connected to two forms of long-term knowledge: long-term schematic knowledge, and long-term specific knowledge (Fauconnier, 1997: p.1). As the discourse unfolds and we structure our thoughts, different mental spaces are constructed and modified—these mental spaces are connected to each other by *mapping mechanisms*, of which two kinds of mappings are primarily important: identity mappings and analogy mappings (Fauconnier, 1997: p.2).

2.4 Fauconnier's Mental Space Theory

(Fauconnier, 1985, 1997) proposes a set of principles for the interpretation of utterances and the assignment of situations to the appropriate mental space. According to (Fauconnier, 1997), words and constructions build mental spaces where the asserted situation is held true in that space only. Between the base space and any built space, there is a mapping of the elements found in each space. Semantic and pragmatic phenomena arise as a product of the possible mappings between spaces.

The Mental Space Theory (Fauconnier, 1985, 1997) states that language guides meaning construction directly in context. According to this view, sentences cannot be analysed in isolation from ongoing discourse. In other words, *semantics* (traditionally, the context-independent meaning of a sentence) cannot be meaningfully separated from *pragmatics* (traditionally, the context-dependent meaning of sentences) (Evans and Green, 2006). This is because meaning construction is guided by context and is therefore subject to situation-specific information. The meaning construction relies on some of the mechanisms of *conceptual projection*. Conceptual projection mechanisms like schema induction establish *mappings*. A mapping connects entities in one conceptual region with another (Evans and Green, 2006).

Mental spaces have specific kinds of information contained in regions of conceptual space. Generalised linguistic, pragmatic and cultural strategies are employed for recruiting information for the construction of such spaces. However, due to the real-time dynamic nature of mental spaces, they result in the formation of 'packets' of conceptual structure, which are unique and temporary, and they are constructed specifically to meet the purpose of the ongoing discourse. The principles of mental space formation and the relations or mappings established between mental spaces have the capacity to yield unlimited meanings.

Depending on the context, an utterance can give rise to different scenarios. This is because the mapping operations, between the state of affairs that holds in reality and the states of affairs that are set up in different versions of the scenario, are guided by context. The same utterance may lead to the representation of a number of different interpretations—these interpretations arise from different mappings between the reality and the scenario that is constructed. Thus, meaning relies on the conceptual processes that connect the links between real situations and hypothetical situations, and is not embedded in the words. These processes result in representations that are consistent with, but only partially specified by, the prompts in the linguistic utterance under consideration.

2.5 Construction of a Mental Space

Linguistic expressions represent partial building instructions, according to which mental spaces are constructed. Of course, the actual meaning prompted for by a given sentence is always a function of the discourse context in which it occurs. This sub-section discusses about the components of a mental space and the principles which guide the formation and interlinking of mental spaces. The following sub-subsections shall introduce the components and principles in brief.

2.5.1 Space Builders

Mental spaces are set up by *space builders*, which are linguistic units that either prompt for the construction of a new mental space or shift attention back and forth between previously constructed mental spaces (Fauconnier, 1997; Evans and Green, 2006). A space-builder is a grammatical expression that either opens a new space or shifts focus to an existing space (Fauconnier, 1997). Space builders can be expressions like prepositional phrases (*in 1962, at the store*), adverbs (*probably, possibly*), connectives (*if . . . then . . .; either . . . or . . .*), and subject-verb combinations that are followed by an embedded sentence (*Fred believes* [*Mary likes bananas*], *Mary hopes . . ., Susan states . . .*), to name but a few (Evans and Green, 2006).

Semantic and pragmatic phenomena arise as a product of the possible mappings between spaces. Space builders include a wide range of semantic phenomena corresponding not only to possible worlds in logical semantics but also a variety of other operators, including temporal expressions, fictional situations, negation and disjunction and various cases in quantification (Fauconnier, 1986). Utterances situate events or states in a *base space* (Fauconnier, 1997) and have elements called *space builders* which set up a new space different from the base space but linked to it.

2.5.2 Base Space

A mental space either represents the base space or is constructed relative to a base space; the base space contains default information which is available to the discourse context at present, including background frames of information which are contextually relevant. The base space contains all the elements which are present in the discourse. Thus, the base space forms the base on which the entire cognitive structures corresponding to the interlinked mental spaces in a discourse are grounded.

2.5.3 Elements

Mental spaces contain *elements*, which are either entities constructed on-line or pre-existing entities in the conceptual system (Evans and Green, 2006). Noun phrases (NPs) are the linguistic expressions that represent elements. Thus, linguistic expressions like names (*Elvis, Madonna, James Bond*), descriptions (*the Queen, the Prime Minister, an African elephant*), and pronouns (*she, he, they, it*) are all elements (Evans and Green, 2006). NPs may be interpreted in the form of a definite interpretation or an indefinite interpretation. The NPs which have indefinite interpretation tend to introduce new elements into an ongoing discourse, that is, these NPs introduce elements that have not already been mentioned in the conversation and are hence unfamiliar (Eg: *I've bought a new car*). However, the NPs with definite interpretation presuppose existing knowledge and therefore, they function in the presuppositional mode. They refer to elements that are already familiar in the discourse, and hence, accessible to the speaker and the hearer, that is, these elements are already a part of the ongoing conversation (*The new car crashes into the store*).

2.5.4 Optimisation Principle

In Mental Spaces Theory, the elements which are introduced in the presuppositional mode, spread to neighbouring mental spaces and are said to be propagated. This process of propagation is controlled by the *Optimisation Principle*. This principle allows elements, together with their properties and relations, to spread through the network or lattice of mental spaces, unless the information being propagated is explicitly contradicted by some new information that emerges as the discourse proceeds (Evans and Green, 2006). Based on this principle, mental space configurations build complex structures with a minimum of explicit instructions.

2.5.5 Properties and Relations

In addition to constructing mental spaces and setting up new or existing elements within those spaces, meaning construction also processes information about how the elements contained within mental spaces are related (Evans and Green, 2006). Space builders specify the properties assigned to elements and the relations that hold between elements within a single space.

Existing knowledge structures like frames and idealised cognitive models structure the mental spaces internally and the space builders, the elements introduced into a mental space and the properties and relations are prompted for recruiting the pre-existing knowledge structure

(Evans and Green, 2006). The verb which is central to the utterance establishes the role-relations between the elements in the mental space.

2.5.6 Counterparts and Connectors

Elements within different mental spaces can be linked by *connectors* which set up mappings between the counterpart elements, the counterparts being established on the basis of pragmatic function. One salient type of pragmatic function is *identity*. For example, in the comics "Batman", the billionaire Bruce Wayne disguises himself as the Batman and fights criminals. The pragmatic function relating the entities referred to as *Bruce Wayne* and *Batman* is co-reference or identity. In other words, both expressions refer to the same individual and hence they form a *chain of reference* (Evans and Green, 2006). Elements in different mental spaces that are co-referential (counterparts related by identity) are linked by an *identity connector* (Evans and Green, 2006).

2.5.7 Access Principle

The Access Principle states that an expression which names or describes an element in one mental space can be used to access a counterpart of that element in another mental space. Formally, the Access Principle given by (Fauconnier, 1986, 1997, 2014) can be stated as—

"If two elements *a* and *b* are linked by a connector F(*b* = F(*a*)), then element *b* can be identified by naming, describing, or pointing to, its counterpart *a*."

Expressions referring to a particular counterpart are essentially bi-directional in nature, that is, they can typically provide access to entities in mental spaces in either direction. Therefore, the connectors can 'link upwards' or 'link downwards' between spaces and under such bi-directional linkage scenarios, the connector is said to be open (Evans and Green, 2006).

2.6 Mental Spaces in Discourse

In the previous sub-sections, we have seen the various components and principles based on which a mental space is built up. In this sub-section, we look at how a mental space is built up. For that purpose, let us discuss an example from (Fauconnier, 1997: p.43). Let us suppose that we are engaged in a conversation about Romeo and Juliet, and the following statement is made:

Maybe Romeo is in love with Juliet.

On utterance of the English sentence, our cognitive faculty brings in a frame from our pre-structured background cultural knowledge about 'Love', and the general argument structure of the verb LOVE(*x*, *y*), denoting '*x* loves *y*', is evoked. This verb, occupying the central structure in the utterance, has two roles highlighted (the lover *x* and the loved one *y*). The utterance also brings up the rich default information linked to the idealized cognitive model tied to this frame.

The word '*maybe*' acts as a Space Builder and sets up a *possibility* space relative to the discourse *base space* at that point. The *base space* contains elements ***a*** and ***b*** associated with the

names *Romeo* and *Juliet*—those elements have already been linked to other frames by background knowledge and previous meaning construction in the conversation. Thus, the new sentence sets up the possibility space, and thereby gives rise to the counterparts ***a'*** and ***b'*** corresponding to the elements ***a*** and ***b***, identified by the names *Romeo* and *Juliette.*
This is possible due to the application of the Access Principle stated in section 2.5.7. The new space is given by the frame '*x* in love with *y*', whose roles are filled by the elements ***a'*** and ***b'***. LOVE(*a', b'*) denotes the internal structure added to a mental space M, such that elements ***a'*** and ***b'*** in space M fit the frame introduced by the verb LOVE. In diagrammatic form, this is expressed in the following kind of representation in Fig.1:

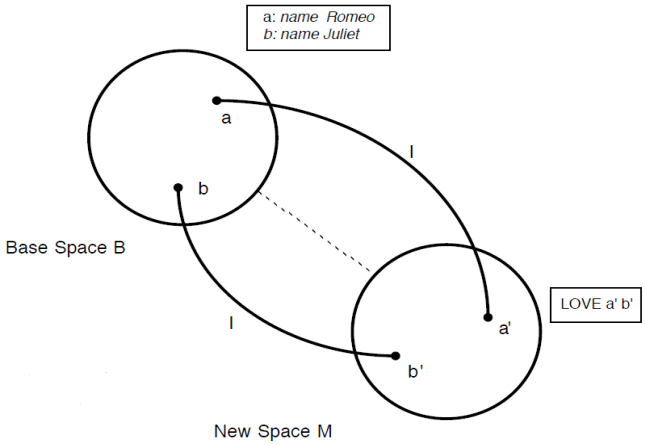


Figure 1: Mappings between Mental Spaces (from Fauconnier, 1997)

A transfer of structure takes place from the parent space to the new space as a default procedure. In the example under study, this creates the effect of associating *a'* and *b'* with the names *Romeo* and *Juliette*, and at the same time, with other background structure for their counterparts *a* and *b* in the base space B.

Similarly, the Mental Space theory can be applied to other discourse types as well. Two or more spaces can blend into a resulting space called *blended space* (Fauconnier & Turner, 2002)—blending is a process of space mapping that pervades human reasoning, and the phenomenon of blending is explored in a wide range of phenomena, most notably metaphor.

2.7 Gärdenfors' Theory of Conceptual Spaces
In Cognitive Science, the two dominating approaches to modeling representations, namely, the symbolic approach and the associationism approach, present competing paradigms to problem solving. However, for cognitive phenomena which cannot be modelled by these approaches, (Gärdenfors, 2000, 2004) advocates a third form of representation using geometrical structures, called conceptual form approach, which is the best suited for describing essential aspects of concept formation. This is the basis of the theory of conceptual spaces.

A conceptual space is built up from geometrical representations based on a number of quality dimensions that often are derived from perceptual mechanisms (Gärdenfors, 2000). The representation of conceptual information is based on implementing geometrical structures instead

of using symbols or connections. Just like physical dimensions form the basis of construction of any object in the real world, the construction of a concept in the cognitive faculty is based on quality dimensions, which are theoretical entities endowed with certain geometrical structures used for modelling the cognitive activities of an organism (Gärdenfors, 2000). An example of such geometrical construct would be the taste tetrahedron in Fig.2, introduced by (Henning, 1916). The four tastes, namely Sweet, Bitter, Sour & Saline, are taken as the four dimensions of taste perception (represented by four planes forming a tetrahedron); a particular taste can be described as a combination of these primary tastes and the resultant taste is mapped onto the planes as a point—however, no taste can exist inside the tetrahedral structure, and is always realized on the surface of the taste tetrahedron. Another example of a conceptual space is the color spindle (see fig.3) which is a double-cone structure based on the Swedish Natural Colour System (NCS) as described in (Sivik and Taft, 1994: p.148).

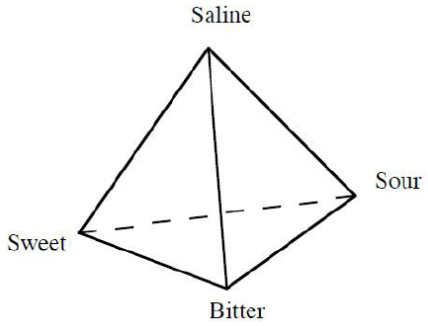


Figure 2: Henning's Taste Tetrahedron (from Henning, 1916)

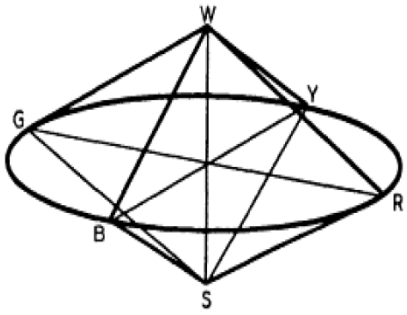


Figure 3: The NCS color spindle (from Sivik & Taft, 1994)

(Gärdenfors, 2000: p.9) also states that:

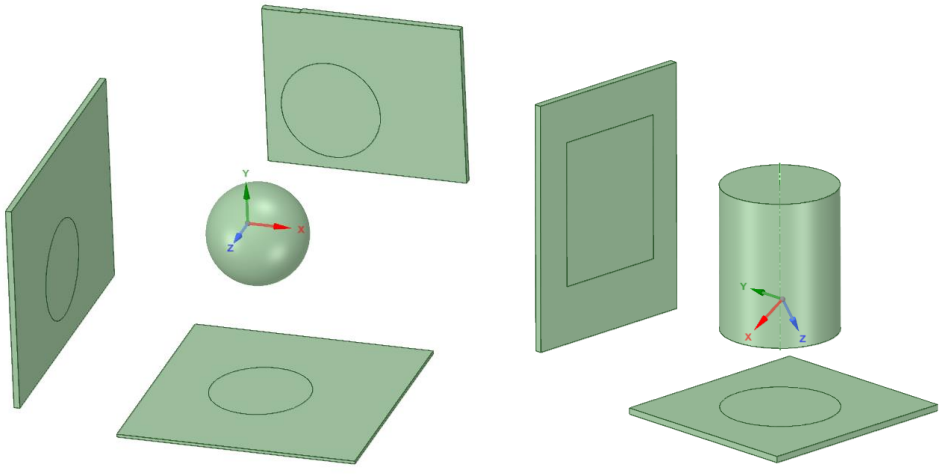
"When the dimensions are seen as cognitive entities, that is, when the goal is to explain natural cognitive processes, their geometrical structure should not be determined by scientific theories that attempt at giving a "realistic" description of the world, but by psychophysical measurements that determine the structure of how our perceptions are represented. Furthermore, when it comes to providing a semantics for a natural language, it is the phenomenal interpretations of the quality dimensions that are in focus."

Further, he also mentions that the human brain contains topographic areas which map sense modalities onto spatial areas. (Quine, 1969) also states that some innate quality dimensions exist in the human brain which are responsible for the learning and development of humans. (Freyd, 1983: p. 193-194) also supported and justified the formation of conceptual spaces in the brain for promoting efficient knowledge sharing through a collaboration between psychological reality and spatial reality. Thus, the theory of conceptual spaces serves as an important modeling approach towards understanding human cognition.

2.8 Projections

According to (Casey, 1893: Ch.11), a projection is the transformation of points and lines in one plane onto another plane by connecting corresponding points on the two planes with parallel lines. If straight lines are drawn from various points on the contour of an object to meet a plane, the object is said to be projected on that plane (Bhatt and Panchal, 2010). The lines meeting the plane on which the object is projected (known as plane of projection), when joined in order, forms a figure which is known as the projection of the object. The corresponding branch of geometry which studies projections is called projective geometry (Veblen and Young, 1918; Hartshorne, 1967). When the projecting lines are parallel to each other and also perpendicular to the plane of projection, the projection is called an orthographic projection (Bhatt and Panchal, 2010). In this paper, in the subsequent sections, the term 'projection' shall refer to 'orthographic projection' only.

Based on the direction from which the projection is initiated or viewed, the plane of projection will change and the projection may be different in different planes. This is demonstrated by two examples—a sphere in Fig.4(a), where the projection obtained is a circle, irrespective of the direction from which the projection is initiated, and in Fig.4(b), a right circular cylinder is shown where the projection varies depending on the viewpoint, that is, the projection is a rectangle when the viewing direction is perpendicular to the axis of the cylinder, but the projection is a circle when the viewing direction is parallel to the axis of the cylinder. However, it may be noted that in case the planes of projection are not orthogonal to the projection lines, but rather oblique (that is, in case of oblique projections), the projected shapes may be distorted with respect to the orthographic projections.



(a) *(b)*

Figure 4: (a) Projections of a sphere onto 3 different planes; (b) Projections of a right circular cylinder onto 2 different planes. (modelled using ANSYS SpaceClaim version 2020 R1 Academic)

[This figure is copyrighted by the author of this paper.]

2.9 Quantifiers

Quantifiers are linguistic expressions that specify or quantify a set. Quantifiers can also be defined as operators prefixed to an open sentence, binding a variable inside it (Dayal, 2012). Examples of quantifiers in English are *every*, *some*, *no*, *few*, *most*, etc. The theory of Generalized Quantifiers is based on an insight of the philosopher Frege that the "quantifiers" can be viewed as relations between two predicates, and this perspective has far-reaching implications for the treatment of quantification in natural language (Lahiri, 2017).

The scope of an operator is the domain within which it has the ability to affect the interpretation of other expressions (Szabolcsi, 1999). Generalized quantifiers have the ability to quantify over properties, and they include proper names, universals, definites, indefinites and complex quantifiers built on them, as well as quantifiers that cannot be captured by first order predicate logic, like *few* and *most* (Dayal, 2012). Scope is understood as a domain in a syntactic notion (Szabolcsi, 1999). Problems of quantifier scope constitute a perennial challenge for uncovering the relations between form and meaning (Ruys & Winter, 2008). The Scope principle states that if two operators govern eachother, they can be interpreted in either scopal order (Reinhart, 1997). This gives rise to scope ambiguity (Kearns, 2011). In case of scope ambiguity, syntactic approaches postulate multiple distinct syntactic representations underlying the same string, with a different meaning assigned to each of them (Ruys & Winter, 2008). The Quantifier Raising (QR) Theory of quantifier scope ambiguity was first proposed by (Chomsky, 1976) and (May, 1977), in which an expression is associated with multiple phrase structure representations—these representations are related by transformations as well as the rules of movement. On this approach, scope ambiguities arise through optionality in the application of a movement rule called Quantifier Raising (QR).

Quantification is a much studied about topic in both logic and natural language semantics. In linguistic semantics, a topic that is currently under investigation in the syntax-semantics interface has to do with the relative scope of quantifiers in sentences that contain multiple quantifiers. It turns out that not all relative scopes are attested with all quantifiers: they are subject to constraints (Lahiri, 2017).

The internal structure of the simplest kind of proposition (atomic proposition) consists of a predicate and its argument/s. A *predicate* describes relationships between entities and is the part that expresses this relationship, whereas the slots containing entities which form a coherent proposition as a part of the predicate's meaning are called *arguments* (Kearns, 2011). The adicity of a predicate is the number of arguments it takes, and thus a predicate can be monadic (one-place), dyadic (two-place), triadic (three-place), etc. Predicates are expressed in a function-argument structure.

The next section shall introduce the hybrid model, under the light of which, the geometric nature of quantifiers shall be discussed in sections 3.1—3.3.

3. Merging Mental & Conceptual Spaces—A Hybrid Approach

In view of the theories presented in the preceding sections, this paper attempts to advocate a synthesis of the two approaches of Mental Space Theory & Conceptual Space Theory into a new hybrid approach, and apply the hybrid approach to understand how quantifiers are perceived and processed in language. At this point, it is important to understand that Quantifiers are just being focused on as a case study in this paper, and the theory may be extended to other elements in

language at a later stage in time. The central idea, which is being described in this paper, is that geometrical structures possessing multiple dimensions are constructed in the form of a mental space, triggered by the linguistic elements which serve as space builders. The hypothesis is presented in the following sub-section.

3.1 Quantifier Space—Adding Geometry to Quantifiers

In any natural language, the set of quantifiers in the language creates a Quantifier Space, which is a mental space for quantifiers. This mental space is effectively a geometrical structure having its extent and geometry defined by the number of quality dimensions it possesses. For quantifier space, the quality dimensions are a manifestation of the number of quantifiers existing in the language under question. Thus, each quantifier acts as a part of a space-builder mechanism and effectively is represented by a plane in this space. Since the number of quantifiers varies depending on the language, the Quantifier Space has a specific geometry which is dependent on the type of the language.

Thus, the Quantifier Space is nothing but a multi-dimensional mental space, the dimensionality being language-specific. As it is true for any geometrical structure, it is also expected that the planes which form the boundary surface of the conceptual space, (that is, the quantifier planes) are oriented at different angles with respect to each other. Together these planes build up a single Quantifier Space, for that language. Thus for a language having 'n' quantifiers, there is an n-dimensional space (called Quantifier Space), which serves as a subspace of the higher dimensional conceptual blending space.

The formation and dissolution of mental spaces in Fauconnier's Theory are dynamic as well as context (and co-text) dependent. Likewise, in the hybrid theory under discussion, the conceptual space belonging to quantifiers is also built up spontaneously as the context demands, but the concept is now generated within and hence enclosed by the generated conceptual space—however, since a geometrical structure is now imparted, with multiple planar boundary surfaces, it is believed that there is a form of activation which is context-sensitive and triggered by linguistic cues. For example, for a language having 'n' quantifiers, the use of a quantifier in an utterance has a two-fold effect: firstly, the n-dimensional quantifier space is evoked within which the central concept is placed (further elaborated in the subsequent paragraphs), and, secondly, the quantifier activates its corresponding plane in the quantifier space while the remaining '(*n-1*)' planes are left inactive.

At this stage, there are two activated units in the mental-conceptual space—the quantifier plane and the concept being generated (which is propositional in nature). The generation of meaning is thus a result of an interaction between these two activated components. The interaction involves the projection of the concept shape onto the activated quantifier plane—therefore, it is expected that quantifier scope is nothing but the orthogonal projection of a 'propositional concept shape' over the quantifier plane which is activated by the speaker under the current context or situation.

The central concept of an utterance is linked to the propositional content being conveyed, and at the heart of any proposition lies the verb which dictates the number of arguments

accepted. Based on the adicity of the verb, a particular concept obtains a specific shape in the mental space. This conceptual shape is a higher-dimensional entity and it depends on how the elements in the language interact. e.g., an intransitive verb gives a different shape to the generated concept in contrast to a transitive or di-transitive verb, taking into account the relational nature of these verbs. For example, the concept shape for the proposition "*John is sleeping*" with monadic predicate SLEEP(*x*) is simpler compared to that for the proposition "*John loves Mary*" with dyadic predicate LOVE(*x*, *y*), which is again simpler with respect to the proposition "*John gave a book to Mary*" with triadic predicate GIVE(*x*, *y*, *z*). This 'apparently arbitrary' multi-dimensional shape is the geometric representation of the '*basic propositional content*' (locutionary act) and lies at the core of the utterance. The meaning obtained from this shape is through a complex series of projections of the shape onto different '*activated*' planes in the mental space of the speaker: which planes are activated depends on the context provided to the proposition.

In case of Quantifiers, the concept shape is produced in the Quantifier Space—the relevant quantifier plane/planes are activated based on the quantifier chosen/intended by the speaker under the given context. The meaning obtained thereafter is a result of the complex interaction between the shape of the concept and the quantifier planes which are activated—in case of multiple quantifiers applied over a single concept, the concept shape is projected onto the resultant plane of the activated quantifier planes, or, it is projected in succession onto the individual component planes to yield the resultant projection. The final resulting projection determines how many elements of the domain or codomain of the function (the verb) are selected as a part of its meaning. Thus, the resulting projection contributes to the semantics.

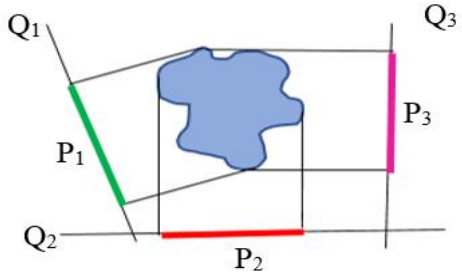


Figure 5: 2-D representation of quantifier planes & conceptual shape of proposition.
[This figure is copyrighted by the author of this paper.]

Since all the generalized quantifiers follow a definite set of properties like Conservativity, Intersectivity, Symmetry, etc. (Lahiri, 2017), it is claimed that a single Quantifier Space exists which includes all the quantifiers as different planes building up the space, with the dimensionality of the space being dependent on the number of quantifiers used in the language. The quantifier planes may have component planes and thus, depending on the quantifier and the prevailing context, a certain plane can be resolved into its component planes.

3.2 Empirical Evidence

As any language has several generalized quantifiers and subsequently a multi-dimensional Quantifier Space, the exact conceptual shape of the proposition is hard to perceive. Also, the complex interactions taking place between a multidimensional entity and its projections in a multidimensional conceptual space are beyond the representational means available to us,

because we are limited to three-dimensional space in our physical world and furthermore on a two-dimensional paper, being limited in the physical dimensions, such representations can never possibly be penned down. Thus, like in case of theoretical physics and in many other sciences and in philosophy, we have to resort to thought experiments (Brown, 1995; Thomason, 1991), particularly of the type of 'mental model accounts' (Andreas, 2011; Bishop, 1998; Cooper, 2005; Gendler, 2004; Palmieri, 2003; Nersessian, 1992, 1993, 2007; McMullin, 1985; Miščević, 1992, 2007). More technical approaches and subsequent analyses to prove this hybrid theory of cognition and language shall be taken up in the future works of the author. Thus, for the present purpose, in order to elaborate on or discuss this complex phenomenon, a crude simplified case is presented as a thought exercise, for an intransitive verb [e.g. HONEST(x)], which is a monadic predicate. Example sentences have been taken as follows:

1. Every man is honest.
2. No man is honest.
3. Not every man is honest.
4. Some men are honest.

Let us take two quantifier planes which represent the quality dimensions of the quantifiers 'Every' & 'No'. The basic proposition 'man is honest' [HONEST(x)] creates a concept shape, which shall get projected onto a Quantifier plane within the Quantifier Space depending upon the activation status of the said Quantifier plane/planes. The projection produced on the plane will apply the properties of the quantifier on the noun phrase projected (say, the properties of 'Every' get applied on the Noun Phrase 'man' which prompts a selection of all the elements in the set of 'man' available in the speaker/hearer's cognitive faculty). Likewise, for the example sentences under discussion, the final output projection of the corresponding interactions shall decide the selection of required number of elements from the set of "man" under the presented context.

For sentence (1), this concept shape when projected onto the plane of 'Every' creates a projection which correlates to the inclusion of all elements of the domain set of "man" under the present context. However, for sentence (2), the same concept shape when projected on the plane of 'No' creates a projection which correlates to the selection of zero elements from the domain set of "man". The correlation mechanism which comes into play shall be studied in further detail in future works of the author. However, it is supposed that the selection of elements may be based on some form of Choice Functions, similar to those discussed in (Reinhart, 1992, 1997), (Kratzer, 1998), (Winter, 1997), (Ruys and Winter, 2008).

For sentence (3), the concept shape is first projected to the plane of 'Every' and $\{\text{projection}_{(\text{every})}\}$ is obtained, which correlates to the inclusion of all elements of the domain set; this projection is now further projected onto the plane of 'Not' to obtain $\{\{\text{projection}_{(\text{every})}\}_{(\text{not})}\}$, which correlates to the exclusion of few elements out of the selected elements, i.e., at least one element is excluded. Thus, the operations may as well be sequential, like relative scope-taking in case of the presence of more than one quantifier in the utterance. From the geometrical viewpoint of this hybrid approach, quantifier scope becomes nothing but the orthogonal projections; rather, more specifically, quantifier scope is defined by the space enclosed or delimited by the projection lines (called projectors) between the concept shape and the projection formed on the plane of projection. The meaning for sentence (4) is obtained in a similar manner like that of the previous examples, and the result is obtained when the concept shape is projected onto the plane of 'Some' to give $\{\text{projection}_{(\text{some})}\}$.

3.3 Scope Ambiguity and Quantifier Raising through the hybrid lens

Using this geometric approach to Quantifiers, a complete treatment of Scope Ambiguity and Quantifier Raising would require a full paper in its own right. However, a brief discussion is presented in this sub-section. Scope ambiguity for multiple generalized quantifiers are being explained in the form of ordering of projections, (i.e., in a sequence of projections, which projection is to be considered for evaluation before another) and the distribution of one projection over another. This is elaborated as follows—for two projections A & B of a conceptual shape onto two different planes, the ambiguity arises from the perspective of ordering and distribution—that is, whether A is taken and distributed over B, or B is taken and distributed over A. Taking the example discussed in (Lahiri, 2017: p.11), for the sentence "*Exactly half the boys kissed some girl*", the scope ambiguity can be explained to arise due to the ordering, whether the projection for 'exactly half the boys' is distributed over the projection for 'some girl', or, the projection for 'some girl' is distributed over the projection of 'exactly half the boys'. An evidence, which may be considered as supporting this claim, is the existence of two logical forms (LF) which allows Quantifier Raising—the relative position of the two NPs at LF will correspond to two distinct interpretations. The two possible LFs for the above example shown in (Lahiri, 2017) are:

a. [S'[exactly half the boys]$_1$ [S'[some girl]$_2$ [S'[S t_1 kissed t_2]]]].
b. [S'[some girl]$_2$ [S'[exactly half the boys]$_1$ [S'[S t_1 kissed t_2]]]].

Thus, there is an ordering of the quantified NPs which gives variation in interpretation. The corresponding interpretations are as follows:

a. Exactly half the boys kissed some girl or other (the other half didn't kiss any girl).
b. There is some (particular) girl, whom exactly half the boys kissed.

According to the presented hybrid approach, the difference in interpretations is due to the difference in ordering of the projections of the concept shape. The hybrid approach also proposes that due to dyadic nature of the predicate KISS(*x*, *y*) in the above example, the concept shape of the proposition incorporates within itself, the component proto-conceptual shapes for the Noun Phrase 'boys' leading to the generation of the projection for 'exactly half the boys', and the proto-conceptual shape for the Noun Phrase 'girl' leading to the generation of the projection for 'some girl'. Thus, the adicity of the verb has a significant influence on the propositional concept shape and its componential structure, which in turn helps to explain the development of scope ambiguity in these cases.

4. Conclusion

This paper presented a synthesis of the Mental Space Theory and the Conceptual Space theory, and laid the foundation for a hybrid approach towards cognitive interpretations of language. The paper also implemented the hybrid approach to quantifiers in natural language, and posited the existence of a multi-dimensional Quantifier Space in the human cognitive faculty. An attempt was also made to explain Scope Ambiguity, Quantifier Raising & relative scopes of quantifiers through the outlined hybrid approach. The geometric understanding of Quantifiers was investigated from the perspective of cognitive linguistics and semantics, along with applications of projective geometry and set-theoretic interpretations. However, it may be noted that this work is only the first of many to come, and future work in this domain shall be directed towards the establishment of this hybrid theory in language cognition on more firm grounds, supported by more rigorous analysis as well as cross-linguistic empirical evidences.

References

Andreas, H. (2011). "*Zur Wissenschaftslogik von Gedankenexperimenten*", *Deutsche Zeitschrift für Philosophie*, 59: 75–91.

ANSYS® SpaceClaim. Version 2020 R1 Academic. ANSYS, Inc. January, 2020.

Bhatt, N. D. & Panchal, V. M. (1960). *Machine Drawing*. Charotar Publishing House Pvt. Ltd. Gujarat, India. Forty-fifth edition, 2010.

Bishop, M. (1998). "*An Epistemological Role for Thought Experiments*", in N. Shanks (ed.), *Idealization IX: Idealization in Contemporary Physics*, Amsterdam/Atlanta, GA: Rodopoi, 19–33.

Brown, J. R. (1995). *Thought experiments*. Canadian Journal of Philosophy, 25(1), 135-142.

Carey, S. (1991). *Knowledge Acquisition: Enrichment or Conceptual Change?* In S. Carey and R. Gelman (Eds.), The Epigenesis of Mind: Essays on Biology and Cognition (pp. 257-291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Casey, J. (1893). "*Theory of Projections*." Ch. 11 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples*, 2nd ed., rev. enl. Dublin: Hodges, Figgis, & Co., pp. 349-367, 1893.

Chomsky, N. (1976). *Conditions on Rules of Grammar*. Linguistic Analysis vol.2.

Cooper, R. (2005). "*Thought Experiments*", *Metaphilosophy*, 36: 328–347.

Croft, W. & Cruze, D.A. (2004). *Cognitive Linguistics*. Cambridge University Press.

Dayal, V. (2012). *The Syntax of Scope and Quantification*. The Cambridge Handbook of Generative Syntax, Rutgers University.

Evans, V. and Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh University Press Ltd. ISBN 0 7486 1832 5 (paperback). Edinburgh, 2006.

Fauconnier, G. (1985) *Mental Spaces*. 2nd Edition. Cambridge: Cambridge University Press.

Fauconnier, G. (1986) *Quantification, roles and domains*. Versus 44/45. 1986.

Fauconnier, G. (1997) *Mappings in thought and language*. Cambridge: Cambridge University Press.

Fauconnier, G. & Turner, M. (2002). *The way we think*. New York: Basic Books.

Fauconnier, G. (2014). *Mental spaces, language modalities, and conceptual integration*. The New Psychology of Language: Cognitive and Functional Approaches to Language Structure (I), 230-258.

Freyd, J. (1983), "Shareability: the social psychology of epistemology," Cognitive Science 7, 191-210.

Gärdenfors, P. (2000). *Conceptual Spaces-The Geometry of Thought*. 2000. Google Scholar Google Scholar Digital Library Digital Library.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought.* MIT press.

Gendler, T. S. (2004). "*Thought Experiments Rethought— and Reperceived*", *Philosophy of Science*, 71: 1152–1164.

Hartshorne, R. (1967). *Foundations of projective geometry*. 1967.

Henning, H. (1916), "*Die Qualitätenriehe des Geschmacks*," Zeitschrift für Psychologie und Physiologie der Sinnesorgane 74, 203-219.

Kearns, K. (2011). *Semantics*. 2nd Edition. Palgrave Macmillan, UK.

Kratzer, A. (1998). "*Scope or pseudoscope? Are there wide scope indefinites?*" In Rothstein, S., editor, Events and Grammar. Kluwer, Dordrecht.

Lahiri, U. (2017). *Generalized Quantifiers in Logic and Natural Language.* The EFL Journal. January, 2017. The English and Foreign Languages University, India.

Margolis, E. & Lawrence, S. (2012). *"Concepts"*. Stanford Encyclopedia of Philosophy. Metaphysics Research Lab at Stanford University. Retrieved 6 November 2012.

May, R. (1977). *The Grammar of Quantification*. PhD Diss., MIT, Cambridge, USA.

McMullin, E. (1985). "*Galilean Idealization*", *Studies in History and Philosophy of Science*, 16: 247–273.

Miščević, N. (1992). "*Mental Models and Thought Experiments*", *International Studies in the Philosophy of Science*, 6: 215–226.

Miščević, N. (2007). "*Modelling Intuitions and Thought Experiments*", *Croatian Journal of Philosophy*, VII: 181–214.

Murphy, G. (2002). *The Big Book of Concepts*. Massachusetts Institute of Technology. ISBN 978-0-262-13409-5.

Nersessian, N. (1992). "*How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science*", in R. Giere (ed.), Cognitive Models of Science, Minneapolis: University of Minnesota Press, 3–44.

Nersessian, N. (1993). "*In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling*", Proceedings of the Philosophy of Science Association, 2: 291–301.

Nersessian, N. (2007). "*Thought Experiments as Mental Modelling: Empiricism without Logic*", Croatian Journal of Philosophy, VII: 125–161.

Palmieri, P. (2003). "*Mental Models in Galileo's Early Mathematization of Nature*", *Studies in History and Philosophy of Science*, 34: 229–264.

Quine, W. V. O. (1969). "*Natural kinds*" in Ontological Relativity and Other Essays, Columbia University Press, New York, NY, 114–138.

Reinhart, T. (1992). "*Wh-in-situ: an apparent paradox*". In Proceedings of the 8th Amsterdam Colloquium. University of Amsterdam, Institute for Logic, Language and Computation.

Reinhart, T. (1997). "*Quantifier Scope: How Labour is Divided between QR and Choice Functions*", Linguistics and Philosophy 20, 335-397.

Ruys, E.G. & Winter, Y. (2008). *Quantifier Scope in Formal Linguistics*. Handbook of Philosophical Logic, 2nd Edition.

Sivik, L., and Taft, C. (1994). "*Color naming: a mapping in the NCS of common color terms.*" Scandinavian Journal of Psychology 35: 144-164.

Szabolcsi, A. (1999). *The Syntax of Scope*. Handbook of Contemporary Syntactic Theory, Baltin-Collins, January 1999.

Thomason, S. G. (1991). *Thought experiments in linguistics*. Thought experiments in science and philosophy, 247-257.

Veblen, O., & Young, J. W. (1918). *Projective geometry* (Vol. 2). Blaisdell.

Winter, Y. (1997). "*Choice functions and the scopal semantics of indefinites*". Linguistics and Philosophy, 20:399-467.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

In Search of Rhythm: A Correlational Study between different Emotive Poetry pieces and their corresponding preliminary Emotional Categorization using Fractal Analytics - A pilot study.

Ratul Ghosh, Shankha Sanyal, Samir Karmakar, Dipak Ghosh
Jadavpur University, India

ARTICLE INFO

Article history:
Received 06/07/2020
Accepted 06/08/2020

Keywords:
Recitation,
Emotion,
Cognitive liguistics,
Psycholinguistics,
Hurst Exponent,
DFA

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

The key objective of the following investigation is to attempt at a preliminary categorization of emotions by subjecting voice recordings of selected poetry pieces (audio signals) to fractal analytics(more specifically DFA). The aim is to calculate Long Range Temporal Correlations (LRTC) that is present in each of the selected audio samples/voice recordings or recitations with the assumption that the scaling exponent (Hurst Exponent) in each part of the audio signal is a reflection of the amount of complexity of the emotional attributes imbued in the same. Furthermore, in the next phase of the investigation, we aspire to use aforementioned audio clips as audio stimuli on participants(around 100) and record their EEG signals and subject the same to biosensor analysis from which we wish to see how the emotional appraisal varies across the different poems chosen.

1. Introduction

Rhythm is to Life as Photons are to Light. Our central nervous system including our brain is composed of tens of billions of neurons, each connected with their neighbouring neurons into several millions of neuronal circuits that we may consider as tiny rhythm machines. And such comparisons are not totally unjustified if we consider how some very specific electrical patterns are responsible for, or can account for, practically all the activities performed by mobile/animate living organisms on earth, from coordinated motion in different species(walking, running, climbing, crawling, jumping, swimming, flying, etc) to the crazy complex circuits that are

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Ratul Ghosh
Email: ratulg.sll.rs@jadavpuruniversity.in

dedicated to emotions such as love, affinity towards abstractions and mathematics or even the human evolutionary biological predisposition to communicate via language. There are millions of neurons and neuronal circuits that are syncing up all the time in our brain, for maybe just a fraction of a second, that generates electrical signals that can be recorded with the use of EEG/EMG(recording magnetic fields generated naturally in the human brain using highly sensitive magnetometers). And these have multiple applications and utilities in various fields. For instance, in the domain of medicine, EEG has many useful applications. One can diagnose and differentiate an unhealthy brain from a healthy one by doing a comparative study of the two. It has been observed that in healthy brains, the waves recorded by EEG are consistent for eg. the beta rhythm during wakefulness while it is the slower delta rhythm during deep sleep. Unhealthy brains, on the other hand, would exhibit abnormal/inconsistent waves which not only would be indicative of something awry with the brain but would also point to the aforementioned and hithertofore enquired and detailed, vital point of correlation between the consistency or rhythm of neuronal signals and their actual role in the human brain. While some have contended that the rhythmic brain waves have a pivotal role to play in our long term memory, to them being ascribed the responsibility to hold information like phone numbers; from dictating the fundamental function of registering our consciousness to the unimaginably crucial task of keeping the different parts of the brain sufficiently synchronised so as to enable the neurons/neuronal circuits from one part of the brain to send strong, unadulterated and unambiguous signals to the other parts or vice versa.

From being responsible for the most essential of living activities (such as movement) to complex activities (such as music, mathematics, language, etc); from experiences to emotions; from registering the state of consciousness to the plausibility of coordinating different parts of the brain, the possibilities of rhythm in the brain is practically endless.

As such our modest endeavour to perform a pilot study to attempt to capture the inexplicably complex correlation between the rhythm in the brain/brain waves (with the use of EEG and fractal analysis thereof) with emotion generated on account of reading a piece of emotive poetry is our very own humble attempt to understand how (or what) the electrical neuronal rhythm in the brain actually accomplishes.

1.1 Broad areas of investigation

Before we move on with the methodology of our proposed pilot study there are a few fundamentals that we must discuss beforehand, in view of their immense importance in our research endeavours.

We shall begin with a brief discussion of brain waves, their subcategorization and what does each type of brainwave have to offer as explanation to the effect of the human brain's biological constitution and function.

Types of brain waves/rhythms :

- Delta wave/rhythm – (0.5 – 3 Hz); site of origin - either thalamus or cortex; associated with NREM sleep (stage 3 deep sleep)

- Theta wave/rhythm – (4 – 7 Hz); site of origin - hippocampus(hippocampal theta rhythm) & cortex(cortical theta rhythm); its function is contentious with Green and Arduini[7] suggesting it to be related with arousal, Vanderwolf[7] with motor behaviour, John O'Keefe[7] with keeping location with environment while Hasselmo[7] points to its role in learning and memory.
- Alpha wave/rhythm – (7 – 15 Hz); site of origin - occipital lobe; associated with wakeful relaxation with closed eyes.
- Mu wave/rhythm – (7.5 – 12.5 Hz); site of origin - primary motor cortex; is observed to be suppressed when one engages in motor activities, most prominent when one is physically at rest. It is held by some that the mirror neuron system suppresses mu wave/rhythm.
- SMR wave/rhythm – (12.5 – 15.5 Hz); site of origin - sensorimotor cortex; associated with states of immobility.
- Beta wave/rhythm [5] – (15 – 30 Hz); site of origin - (possibly) thalamus, deep within the brain; associated with normal waking consciousness linked to active concentration or anxious thinking.
- Gamma wave/rhythm [6] – (>30 Hz); site of origin - observed at cortical as well as subcortical areas of the brain; there is some disagreement over the role/association of gamma waves with plausible brain functions with some contending that it has a role to play with consciousness while others oppose the view.

From the above discussion, we shall direct our major study to investigations of theta, alpha, mu, smr and beta waves/rhythms for all practical purposes.

But what do we mean by study or some particular brainwave/rhythm. Since brain waves are, by their very nature, non-linear, we shall have to employ a standard of analysis that can meet the challenging complexity of non-linearity. This is where fractal analytics steps in.

Fractals has been defined within the ambit of mathematics as the subset of Euclidean geometric space where "the fractal dimension exceeds the topological dimension" [1] . By fractal analysis we mean the assignment of fractal dimension and other fractal characteristics to our electrical signal extracted from the EEGs performed on multiple volunteers [1,2].

In the present pilot study, we will restrict ourselves to the preliminary emotional categorization of selected poetry pieces using robust fractal tools on the acoustic signals.

In order to do so, we would need to incorporate such tools of analysis that would enable us to tackle the nonlinear dynamical nature of the input signals. We would need to incorporate the mathematics of chaos, fractals and power laws in order to accomodate for such analysis. But before we venture any further, we feel it to be prudent to mention or attempt to introduce what the nature of such tools are with brief lucidity.

2. Description of Chaos

Let us begin with Chaos. Chaos has been variously described by scholars and mathematicians for quite some time now.

Some have contended that it is "A kind of order without periodicity", while others have claimed "Chaos is apparently random behaviour with purely deterministic causes"[11].

I, personally, perceive chaos as "Man's attempt to grapple with the dynamic complexities in Nature."

For instance, we can take the example of the famous "Butterfly Effect".
The Butterfly Effect asks questions such as : "Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas ???"
The quest to find answer to such enquiries beckons the arrival of Chaos to the floor to attempt at resolutions.Chaos attempts to tackle such mind bogglingly cryptic riddles, some of which, admittedly, come off as utterly ludicrous at first sight. And it does so on the premise of the reality of self similarity/ existence of self similar systems [9, 10].

2.1 Fractals & Self Similarity

Self-similarity means that a structure, or a process, and a part of it appear to be the same when compared. And our world is filled us such self similar systems (some of immense complexity)

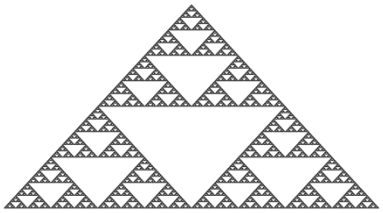
For instance: coastlines, faults, joint systems, Folds , topographic features, turbulent water flows, drainage patterns, Clouds, trees, leaves, DNA, bacteria cultures, Broccoli, lungs, heart, brain. Some even go as far as to contend that perhaps the entire Universe is a self similar system when viewed from the proper perspective [1, 2, 9].

To deal with such self similar systems we incorporate the geometric concept of fractals.

In mathematics, a fractal is a subset of a Euclidean space for which the fractal dimension strictly exceeds the topological dimension [1, 7].

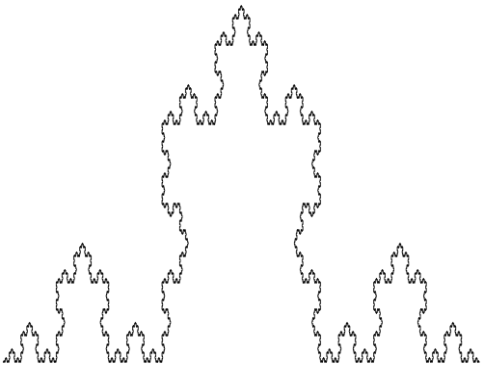
Or, in other words, a fractal is a geometrical pattern that is iterated at even smaller or larger scales to produce self similar irregular shapes or surfaces that cannot be represented by Euclidean geometry [1].

A few diagrammatic/illustrative examples of fractals will perhaps render greater clarity [7].

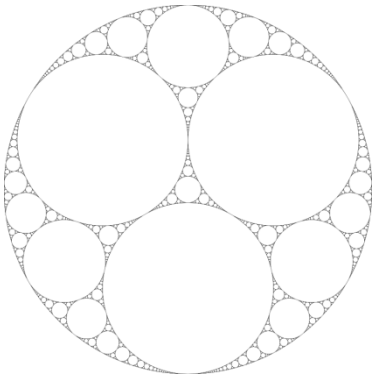


1.The Sierpinski triangle [7, 4]

2.The Koch Curve [7, 4]



3.The Apollonian gasket [7]



2.2 Detrended Fluctuation Analysis (DFA)

DFA stands for **Detrended Fluctuation Analysis** [1, 2]**.**

It is an algorithmic tool for scaling analysis which is used to estimate long-range temporal correlations of power-law form (Peng et al., 1995).

So, if there is a sequence of events that exhibits a non-random temporal structure which shows slowly decaying auto-correlations, we can use DFA to quantify how slowly these correlations decay as indexed by the DFA power-law exponent.

Typical classes of correlations are reflected in DFA exponents 'α':

Uncorrelated sequence: $\alpha \sim 0.5$
Anti-correlated sequence: $0 < \alpha < 0.5$
Long-range temporal correlations: $0.5 < \alpha < 1$
Strong correlations that are not of a power-law form: $\alpha > 1$

3. Methodology

At the pilot stage of this experiment we shall perform fractal analysis of selected poetry pieces and endeavour to achieve / attempt at a preliminary categorization of emotions (emotional categorization) of aforementioned poetry pieces.

For the purpose of our pilot study we have adopted the following methodology for designing our experiment.

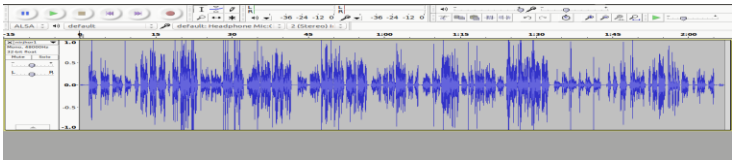
We have selected 4 poems by Rabindranath Tagore viz,

1. Nirjhorer Shopnobhongo(নির্ঝরের স্বপ্নভঙ্গ)
2. Africa (আফ্রিকা)
3. Shah Jahan (শাহ জাহান)
4. Puroshkar (পুরস্কার)

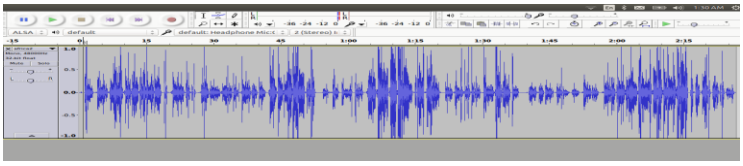
All the audio signals chosen were of two minutes duration each.

The corresponding acoustic wave representation/profile are as follows:

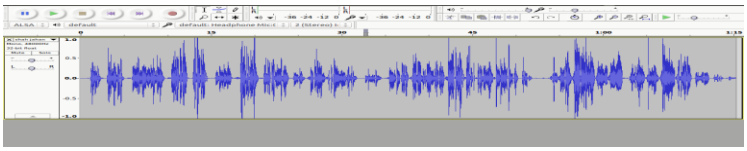
Nirjhorer Shapno Bhanga (নির্ঝরের স্বপ্নভঙ্গ):



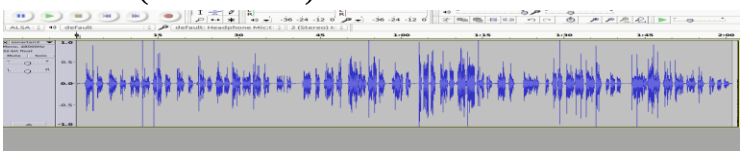
Africa (আফ্রিকা):



Shah Jahan (শাহ জাহান):



Sonar tori (সোনার তরী):



We have done the following:

- The two minute clips of the four audio samples were subjected to DFA analysis in four parts considering 30 secs clips.
- The DFA scaling exponent computes the amount of long range temporal correlations present in each of the audio samples.
- The scaling exponent in each part is a reflection of the complexity of the acoustic signal and it is expected that this complexity will be reflection of the emotional attributes present in the recited piece.
- The amount of LRTC (Long Range Temporal Correlations) is a measure of HE (Hurst Exponent).

At the pilot stage of this experiment we shall perform fractal analysis of selected poetry pieces and endeavor to achieve / attempt at a preliminary categorization of emotions in the aforementioned literary pieces.

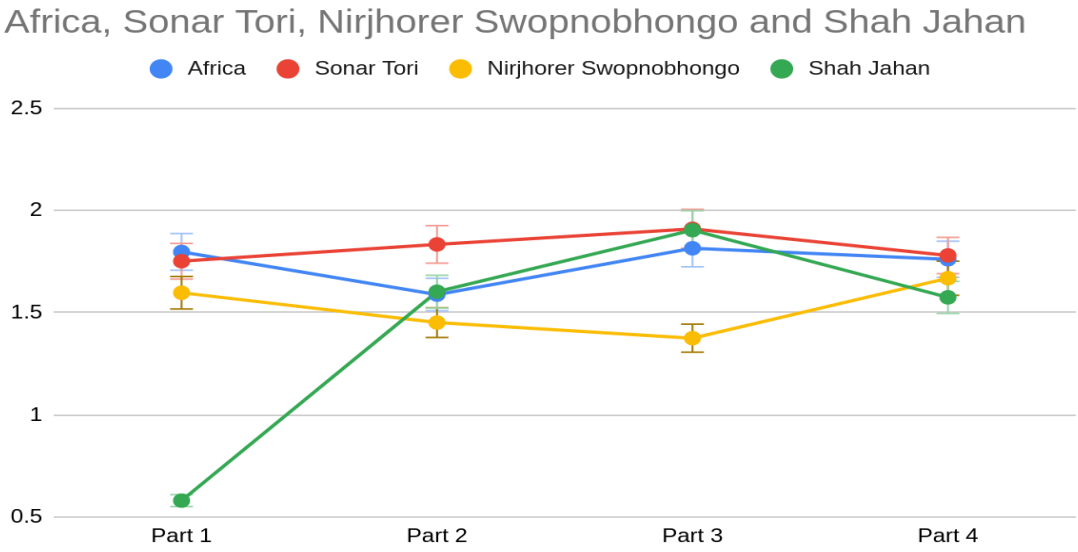
5. Observations

We made the following observations.

DFA Scaling Exponent

	A	B	C	D	E
1		Africa	Sonar Tori	Nirjhorer Swopnobhongo	Shah Jahan
2	Part 1	1.796861731	1.751438024	1.597212168	0.5800755775
3	Part 2	1.58805573	1.833661866	1.451260895	1.602101351
4	Part 3	1.81475374	1.910105144	1.374788218	1.903503613
5	Part 4	1.761040418	1.779717045	1.668331941	1.574686276

And on plotting the same we obtained the following graphical representations:



6. Discussion

In case of the poem -

- *Sonar Tori, the variation of scaling exponent is found to be minimal.*
- *The variation of the scaling exponent is maximum for Shah Jahan(Taj Mahal) and Nirjorer Shopnobhongo.*
- *In the case of Africa and Nirjhorer Shopnobhongo, there is a distribution of higher and lower values of scaling exponent.*
- *In some portions of the recitations for eg. Part 3 of Africa, Sonar Tori and Shah Jahan, the Hurst Exponent values are found to be very close to each other. This may be due to the fact that emotional manifestations in these portions are very close to each other. This could be verified with the help of a human response study where participants would be made to rate these clips on the basis of their emotional arousal.*
- *The poem Nirjhorer Swopnovongo maintains a distinct level of scaling exponent, quite different from others. This indicates that the emotional appraisal for this particular poem is largely different from the other two.*

7. Future Plans

We have decided to take the aforementioned four audio clips of recitation recordings and use it as audio stimuli to L1 Bengali speakers while we record their corresponding EEGs. The audio stimuli will be delivered via pre-recording the aforementioned poems. The poems(or selected sections/portions thereof) will be recited emotively so as to recreate the same emotion in the hearer/volunteer. But in the pilot study we shall restrict ourselves to performing fractal analysis to develop a preliminary emotional categorization of the selected poetry pieces.

The ultimate purpose of this experimental exercise would be to find correlation, if any, between the emotion evoked in the brain of the hearer due to the audio stimuli and the corresponding manifestations observed in the EEG signals thus recorded and subsequently subjected to fractal analytics(on account of their inherent natural non-linearity).

This pilot study has the potential to unravel not only the correlation between emotion evocation in humans via audio stimuli but also to provide solutions or future directions for us to further our investigations to understand the underlying neurophysiology as well as the neuroelectrical signal manifestation of cognition of emotion.

It shall further shed light on how emotion is neurocognitively cognised in L1 speakers when exposed to emotive audio stimuli in the same L1 language- thereby providing us with empirical data for further research, analysis and reinterpretation from a purely neurolinguistic angle.

In other words we will perform:

- Listening test on a sample of 100 respondents to see how the emotional appraisal varies across the different poems chosen.

- Biosensor analysis using EEG data for which the different chosen parts will be given as auditory stimuli.

- EEG signal processing and analysis will be done.

8. References

1. Sanyal, Shakha; "Chaos based Neuro-Cognitive Physics study of Complex System(Music) on the Human Brain"; Ph.D. thesis, Department of Physics; Sir C. V. Raman Centre for Physics & Music, Jadavpur University, India; 2018.
2. Ghosh, Dipak; Sengupta, Ranjan, Sanyal, Shankha; Banerjee, Archi; "Musicality of the Human Brain through Fractal Analytics"; Springer, 2018.
3. Alligood, T., Kathleen, et al; "Chaos : An Introduction to Dynamical Systems"; Springer, 2000.
4. Strogatz, H., Steven; "Nonlinear Dynamics and Chaos"; Taylor & Francis, Reprint 2018.
5. Maxwell, Sherman; et al; Proceeding of the National Academy of Sciences, USA; article titled, "New theory explains how beta waves arise in the brain"; Brown University website. Link : h ttps://www.brown.edu/news/2016-07-25/beta , 2016.
6. Jia, Xiaoxuan; et al; "Gamma Rhythms in the Brain"; PLOS Biology, PMC, US National Library of Medicine, National Institutes of Health; NCBI online resources; Link : h ttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3084194/ ; April, 2011.
7. Wikipedia online articles on brain waves, types of brain waves etc.
8. Dipak Ghosh, Shukla Samanta, et al.; "Multifractals and Chronic Diseases of the Central Nervous System", Springer, 2019.
9. Schroder, Manfred; "Fractals, Chaos, Power Laws"; Dover Publications Inc.; 1991.
10. Gleick, James; "Chaos"; Viking Books; 1987.
11. Argyris, J., Faust, G., Haase, M., Friedrich, R.;"An Exploration of Dynamical Systems and Chaos"; Springer; 2015.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Development and standardisation of Bengali sentence identification test

Mousumi Chatterjee,[1] Indranil Chatterjee,[2] Palash Dutta,[2] Krishna Kali Banerjee[2]
[1]*iHEAR*, [2]AYJNISHD(D), RC, India

ARTICLE INFO

Article history:
Received 15/05/2020
Accepted 06/08/2020

Keywords:
speech perception, native language, background noise, communication, standardisation

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

In recent years verbal communication occurs mostly in intrusive noisy situation. Therefore it is important to judge the speech perception ability of individual. This necessitates the development of speech perception tests. A wide range of speech perception tests are available in different language. No such test material is accessible in Bengali language. Hence this study aimed to develop and standardise the speech perception test in Bengali language

A total 600 Bengali sentences were selected from newspapers, magazines, daily conversation. These sentences were subjected to test for naturalness and predictability. Subjects' task was to rate the sentences in a five point rating scale where 5= natural, 4=not so natural, 3= doubtful, 2= un-natural, and 1= artificial. Sentences that were rated 4 or more than 4 in naturalness testing and 2 or more than two in predictability were included. Predictability of the sentence was assessed by identification of key words. Key word in each sentence varied from 2-4 words. A sentence that scored 2 or more than 2 in predicting the missing target word was integrated in the list. A total of 314 sentences were selected for further utilization The Bengali sentence identification test revealed a normative identification score of 51.70% at global SNR (-5 dB). The Bengali sentence identification test developed can be used to evaluate the therapeutic progress, hearing aid performance, hearing aid selection and most importantly to assess the speech perception ability.

1. Introduction

Human beings use speech as a means of communication. This use of spoken language makes us unique in the animal kingdom (Fitch, Hauser, & Chomsky, 2005). Speech signal undergoes various changes related to its frequency, intensity and temporal characteristics. Listener need to carefully identify all this changes of a signal to correctly comprehend the meaning. This is done by the process of speech perception. This ability of the speech perception depends on the neural integrity of the auditory system and the redundancy of acoustic information in spoken language (McGuffin, 2007). Speech is perceived

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Krishnakali Banerjee
Email: krishnakalibanerjee96@gmail.com

following detection, discrimination, reorganization and comprehension (Holt and Lotto, 2010). speech perception ability of client relative to the pure tone or other stimuli.

A dynamic change occurs in speech signal during conversation, which includes continuous discourse. This change results from alteration of fundamental frequency, intensity, speed, articulation, emotional emphasis, voice inflection, distance from speaker to listener and eternal juxtaposition of words and their phonemic elements. In recent years verbal communication occurs mostly in intrusive noisy situation. This background noise affects the acoustical properties of speech. Listener's sensory, cognitive and neuronal resources should be strong enough to extract the speech signal in adverse condition and achieve successful processing of speech (Wong, Uppunda, Parrish, and Dhar, 2008). It is observed that speech perception ability deteriorates as the perception environment changes. Speech perception in noise is an extremely important and challenging task and requires specialised auditory tasks like acoustic analysis, auditory-motor integration, phonological memory as well as auditory attention. In realistic acoustical environment we hear various sounds but we listen to a particular sound from the muddle of all. We focus our attention to the particular sound of intension. This trend is called "cocktail party effect" (Cherry, 1953). The process of speech perception starts from detection of signal and proceeds by evaluating it in different dimensions, segregating it from background noise, and attending, recognizing and finally comprehending the meaning. All this functions are performed in the brain (Musiek & Chermak, 1994).

2. NEED OF THE STUDY

As verbal communication occurs mostly in intrusive noisy situation, it is important to judge the speech perception ability of an individual. Speech perception ability is highly influenced by the language. Human individual shows better speech perception ability in their mother tongue than other languages (Delattre, 1964). Native listener can better identify the phonemes of native language than non native listeners. This ability is also present in noisy environment (Jin and Liu, 2014). Liu and Jin (2015) compared the Chinese and Korean vowel detection threshold in Chinese and Korean native listeners. Their study reveals that individuals were a better performer in their native language for speech stimuli identification where both the group were equally good in non speech sound identification. A wide range of speech perception tests are available in different language. No such test material is accessible in Bengali language. Hence this study aimed to develop and standardise the speech perception test in Bengali language.

3. METHODOLOGY

Total 93 participants were included in various phase of the study of which 33 adult subjects were included in different steps of sentence list development that is by estimating naturalness, predictability and sentence equivalency and 60 adults were engaged to standardise the sentence list. . All the participants had normal hearing sensitivity (hearing threshold $\leq$ 25 dB HL) and age range was between 18-50 years. Mean age was 35.06 years. The study was conducted in two phase.

PHASE I

Development of sentence material

A set of 600 sentences were selected from the daily conversation, newspaper, and magazine. Sentences were selected base on following criteria:

1) The total number of words ranged from three to five

2) The number of syllable was 8 to 9 in a sentence.

3) Sentences were familiar and contained equally difficult words.

4) No words in a sentence contained more than three syllables.
5) Conversational speech was included in the sentences.
6) Proverbs, exclamations, questions, proper name was excluded.
7) Sentences were syntactically correct.
Linguistic verification of the sentences was done based on these criteria.
Naturalness and Predictability testing
A total of 600 sentences were selected based on the above mentioned criteria. These sentences were further endured for naturalness and predictability testing. For naturalness testing, 8 subjects with normal hearing were asked to rate the naturalness of the sentences in a five point rating scale (5= natural and 1=artificial). Sentences rated more than four were included in the list. Semantic naturalness of the sentences and their occurrence in normal conversation were considered for naturalness testing. For predictability testing sentences were presented to same 8 normal hearing subjects. Predictability was assessed by the identification of key words. Key words are those words of a sentence which are supposed to be essential for comprehension of the sentence. All the sentences were presented with the missing target words and participants were asked to presume the words that might occur. Sentences that were rated more than two were included in the list. Likewise a total of 314 sentences were considered for steps forward.

Sentence recording and editing
Total 314 selected sentences that were considered as natural and predictable was recorded in a sound treated room. A microphone was placed at a distance of 5cm from the speaker's mouth. Recording was done at a sampling frequency of 44,100Hz and 16 bit analog to digital converter was used to digitize the signal. Wave surfer software version1.8.8p3 was used for recording the sentence material.
The sentences were spoken by a native young adult female speaker of Bengali language. Sentences were developed in the standard Bengali dialect. The speaker was instructed to retain a stable vocal effort, clear articulation and the native intonation pattern throughout each sentence. Adobe Audition software (version 1.5) was used to edit the recorded sentences. Editing was done for removal of breathing noise or any other noise. Silent interval at the beginning and ending of the sentences was also edited.

Noise generation and mixing
Recorded sentences were spectrally analysed and the long term average speech spectrum (LTASS) was derived. For noise generation the sentences were concatenated in random order and Fast Fourier Transformer (FFT) was performed on this. Then phase of the FFT was randomized and this was subjected to generate the spectrally shaped noise. This whole work was done using MATLAB (MathWorks USA) software. The spectrum of the generated noise corresponds to the spectrum of the sentences.
Speech spectrum shaped noise provides good masking effect than other type of noise (Festen and Plomp, 1990). The psychometric function of the sentences provides maximum slope in presence of spectrum shaped noise and the speech recognition threshold determination is also found to be highly accurate. Therefore all the sentences were mixed with the spectrally shaped noise at different SNR level from -7 to 0. The RMS amplitude of the speech and noise signals were measured in 50 millisecond bins.

Determination of global SNR
Global SNR can be defined as the SNR level at which subjects score 50% on speech identification test. Global SNR is determined to overcome the intrinsic limitation of speech perception resulted from ceiling

and floor effect. In our study an adaptive procedure was used to meet the global SNR. A pilot study was done with 10 native Bengali speaking subjects aged from 18-50 years. All the sentences were delivered monaurally to each participant at each SNR level in 1 dB increment. Subjects were instructed to repeat the sentences they can hear. From their responses correct identification of key word in each sentence was attained. The outcomes of the study shows that subjects were scored approximately 50% correct identification of key words at -5 dB SNR. Therefore -5 dB SNR was considered as global SNR.

Preparation of an equivalent subset of testing

Intelligibility of the sentences may not be equal when presented with a spectrally shaped noise. There are some other factors that also may influence the perception of sentences like, word familiarity, intonation variations, and intensity variations. For this reason sentence equivalency assessment was done. This phase was aimed to eliminate the sentences that are too easy or too hard for identification. Sentence equivalency test was performed at three SNR level. One at the global SNR that is -5 dB and other was on either side of the global SNR that is -3 dB and -7 dB SNR. Total 15 native speaker of Bengali aged from 18 to 50 years were incorporated for Sentence equivalency test. Subjects were divided in three specified SNR group having five members on each group. Total 314 sentences were presented on five participants of -5dB SNR, another five participants on -3 dB SNR, a different five participants on -7 dB SNR.
Sentences were played through a Lenovo Core™ i3 laptop and that was connected with a calibrated audiometer (MAICO MA 53). A TDH 39 headphone was used for monaural presentation of the sentences at their most comfortable level.

The participants were asked to repeat the sentences as accurately as possible and their responses were recorded in a recorder. Scoring was done based on identification of key words of the sentence. A score of 1 was given to each correctly identified key word and 0 to incorrect key word identification. Misarticulated, incorrect or partially correct word repetition was scored as 0. Key words in the sentences varied from two to four. Calculation of total correctly repeated key words were done in two methods. First, total number of correctly identified key word for each participant at each SNR was counted. Second, the mean value of correctly identified key word at three levels of SNR was calculated. Then the number of correctly identified key words for each sentence was compared with the mean value of correctly identified key word at -3 dB SNR, -5 dB SNR and -7 dB SNR levels. A sentence with scores above or below the mean was eliminated. With this process total 69 equivalent difficult sentences were shortlisted. Short listed equivalent difficult sentences were further divided into six sentence lists. Each list consist total 10 sentences and remaining 9 sentences were used for prior practice. Special attention was given on equal distribution of all phonemes over the sentence lists.

PHASE II

Standardisation of sentence list

In this phase standardisation of the prepared sentence list was done. 60 participants of age range 18 to 50 years with normal hearing (Hearing threshold $\leq$ 25 dB HL) was included. A routine audiological test was done to confirm the normal middle ear functioning and normal hearing level. Prepared sentence list was presented monaurally to each participant at the global SNR level of -5 dB. Stimulus sentence lists were played from a Lenovo core™ i3 laptop that was connected with the calibrated MAICO MA 53 audiometer. Sentences were presented via TDH 39 headphone.

Subjects were instructed to repeat the sentences they can hear and their response was recorded in a recorder. Scoring was done based on identification of correct key word in each sentence. Before doing the test a practise session was performed for subject's familiarisation of the test.

Statistical analysis

SPSS software (version 1.8.8.p3) and Microsoft excel (2007) software were used to conduct the statistical analysis. Statistics included estimation of mean and standard deviation and percentage of identification score. ANOVA: Single Factor test was done to observe the performance variation across lists, if any. Bonferroni post hoc analysis was done to identify the list that varied significantly. Pearson correlation coefficient was measured for test retest reliability.

4. RESULT

Development of sentence material

600 Bengali sentences were selected from the magazine, news papers and daily conversation. All the sentences contained a word range of three to five, and a syllable range of eight to nine. These sentences were subjected to test for naturalness and predictability. A pilot study was conducted on 8 normal hearing subjects. Subjects' task was to rate the sentences in a five point rating scale where 5= natural, 4=not so natural, 3= doubtful, 2= un-natural, and 1= artificial. A sentence whose mean score was 4 or more than 4 was included in the list. Based on this result a set of 492 sentences were selected for predictability testing. Predictability of the sentence was assessed by identification of key words. Key word in each sentence varied from 2-4 words. A sentence that scored 2 or more than 2 in predicting the missing target word was integrated in the list. After that total 314 sentences were selected for further utilization.

Determination of global SNR

One more important objective of the study was to determine the global SNR at which identification of key words yield a score of 50 %. Global SNR was determined to avoid the ceiling and floor effect of the speech intelligibility. The entire sentences were recorded in Wave surfer software version1.8.8p3. Digitization of the recording was done at a sampling rate of 44,100Hz and 16 bit resolution. A spectral shaped noise was then mixed with this recorded sentence at different SNR level, starting from 0 to – 7 dB SNR using Adobe Audition software version 1.5.

Recorded stimulus was presented to 10 Bengali native subjects aged from 18 to 50 years and they were asked to identify and repeat the sentence they can hear. Stimulus was presented monaurally at their most comfortable level. Scoring was done based on correctly identification of key words. Mean identification of key words at each SNR level was calculated. It was observed that at -5 dB SNR percentage of key word identification score was approximately 50 %. Hence - 5 dB SNR was considered as global SNR.

Sentence equivalency assessment

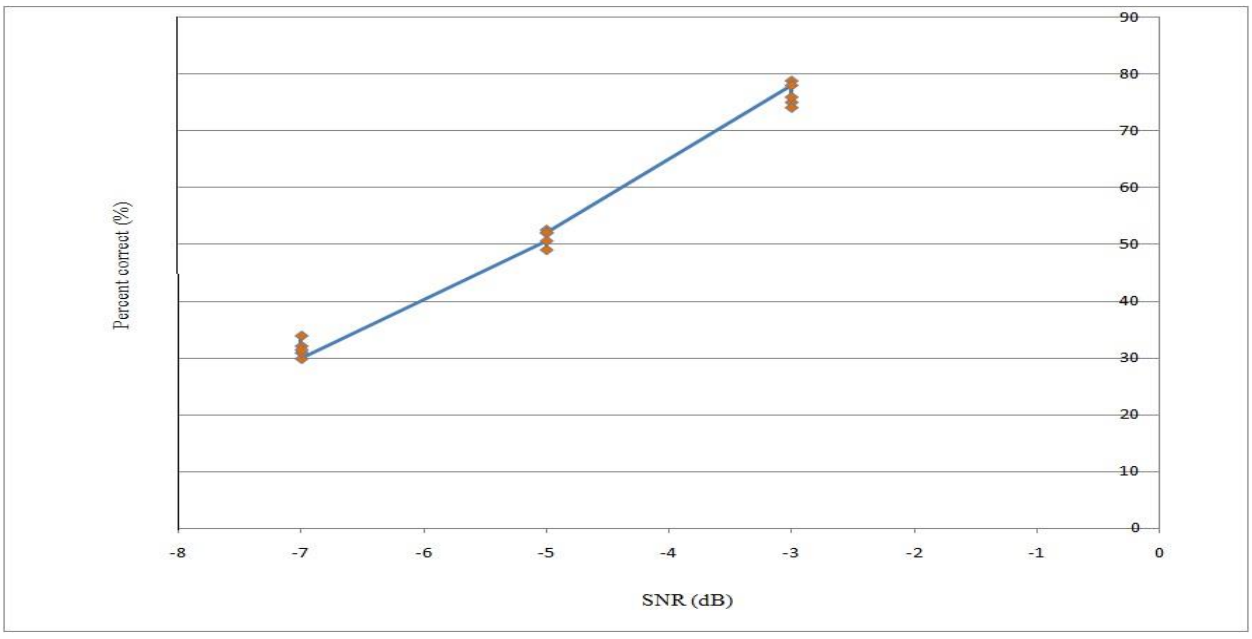
Sentence equivalency assessment was done with 314 final sentences at three SNR level (-3 dB SNR, -5 dB SNR and -7 dB SNR). Sentence equivalency test was done to prepare a set of sentences that have equal difficulty in perception. Results of sentence equivalency assessment was obtained with Microsoft excel (2007) software.

Table 1-Percentage correct, mean and standard deviation of sentence identification score at three SNR group.

Signal to noise ratio (SNR)	-3 dB SNR	-5 dB SNR	-7 dB SNR
Percentage correct	76.27%	51.30%	31.53%
Mean	1.86	1.25	0.77
SD	0.98	1.00	0.77

From the table 1 it can be observed that at -3 dB SNR percentage of correctly identified key word was 76.27% with the mean of 1.86 and standard deviation of 0.98, at -5 dB SNR percentage of correctly identified key word was 51.30% with the mean of 1.25 and standard deviation of 1.00, and at -7 dB SNR percentage of correctly identified key word was 31.53% with the mean of 0.77 and standard deviation of 0.77.

Sentence equivalency test was done to measure the difficulty level of all the sentences. Mean score of correctly identified key word at each sentence was compared with the mean of correctly identified key word at each SNR. The sentences whose mean score was more than SNRs mean score was supposed to be easy and the sentences whose mean score was less than the SNRs mean score was thought to be hard to identify. Based on this 254 sentences eliminated from the list among which 149 was too easy and 105 was too hard to identify.



Graphical representation of the mean key word identification score at three SNRs. -3 dB SNR, -5 dB SNR and -7 dB SNR. Each symbol represents each individual.

Total 69 sentences were finally selected and distributed equally to develop the sentence identification test list. Finally six sentence lists were prepared with ten sentences in each list and remaining nine sentences were used as practice list to familiarize the test procedure.

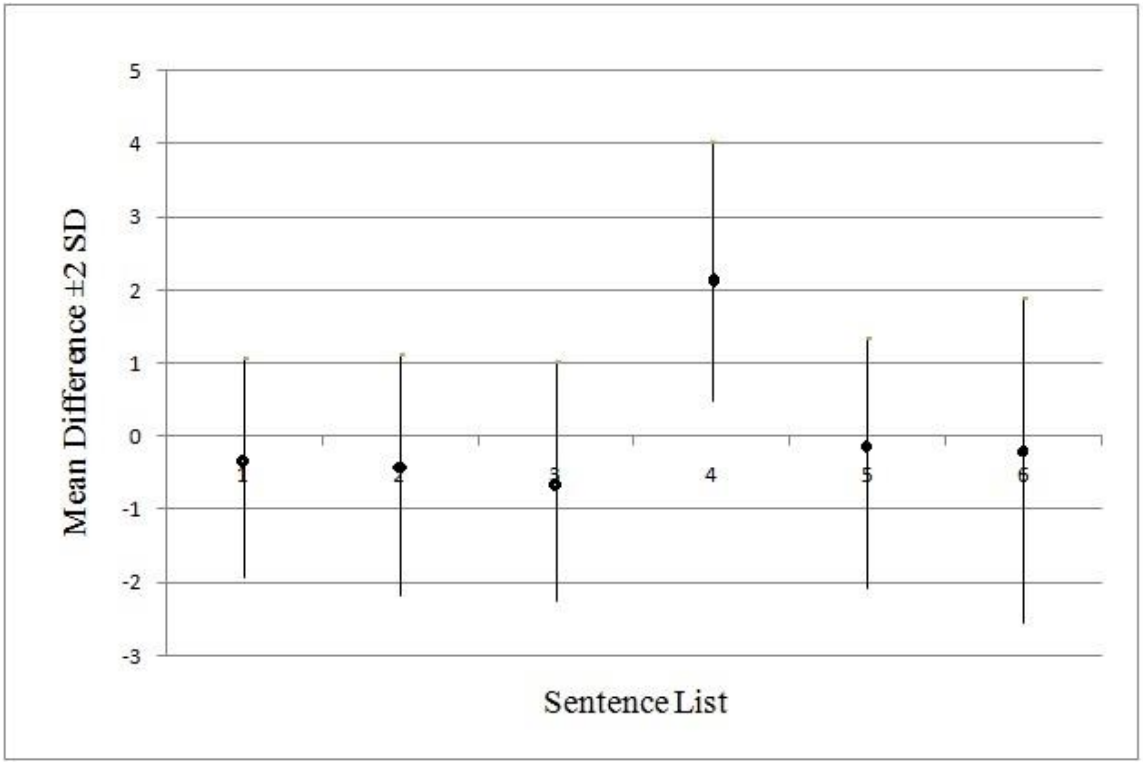
Standardization of test material

Developed sentence material was presented to 60 normal hearing listeners, age range 18 to 50 at -5 dB SNR to achieve a standardize sentence identification score. Stimulus was presented monaurally and scoring of response was done based on identification of key words. Table 2 exemplifies the mean and standard deviation of six lists.

Table 2: Mean score and standard deviation of performance of six lists.

Sentence List	List 1	List 2	List 3	List 4	List 5	List 6
Mean	12.93	13.03	13.11	10.26	12.85	12.7
SD	0.73	0.90	0.84	0.93	0.98	1.34

From table 2 it can be seen that all list revealed an identical performance score. Only the score of list four was quite different from other lists. Hence, mean difference and standard deviation of the performance across list was calculated for all the participants. On this new set of information ANOVA: Single Factor test was applied. Mean difference is the difference of mean score of each participant and the obtained score of each participant at each list. Mean difference and standard deviation of six lists were graphically represented. In each list participants were 60 (n=60). For each list ±2 standard deviation from the mean was considered.



Mean difference and standard deviation of six lists. Error bars represents ±2 SD from the mean.

ANOVA: Single Factor test was done with SPSS software version 16.0 to evaluate that there was no significance performance variation across the lists. ANOVA result shows a significant performance variation across list at – 5 dB global SNR [F =92.961, p<0.05].

Table 3: ANOVA: Single Factor test result revealed a significant (p<0.05) performance variation across the list.

ANOVA					
Performance					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	363.893	5	72.779	92.961	.000
Within Groups	277.145	354	.783		
Total	641.038	359			

Results of ANOVA test revealed that there was a significant difference in performance across list at -5 dB SNR. To determine the specific list that varies in performance, Bonferroni post hoc analysis was done.

Test results of Bonferroni post hoc analysis.

performance Bonferroni						
(I) list	(J) list	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	.11667	.16154	1.000	-.3607	.5941
	3	.20333	.16154	1.000	-.2741	.6807
	4	-2.65000*	.16154	.000	-3.1274	-2.1726
	5	-.05000	.16154	1.000	-.5274	.4274
	6	-.10000	.16154	1.000	-.5774	.3774
2	1	-.11667	.16154	1.000	-.5941	.3607
	3	.08667	.16154	1.000	-.3907	.5641
	4	-2.76667*	.16154	.000	-3.2441	-2.2893
	5	-.16667	.16154	1.000	-.6441	.3107
	6	-.21667	.16154	1.000	-.6941	.2607
3	1	-.20333	.16154	1.000	-.6807	.2741
	2	-.08667	.16154	1.000	-.5641	.3907
	4	-2.85333*	.16154	.000	-3.3307	-2.3759
	5	-.25333	.16154	1.000	-.7307	.2241

	6	-.30333	.16154	.919	-.7807	.1741
4	1	2.65000*	.16154	.000	2.1726	3.1274
	2	2.76667*	.16154	.000	2.2893	3.2441
	3	2.85333*	.16154	.000	2.3759	3.3307
	5	2.60000*	.16154	.000	2.1226	3.0774
	6	2.55000*	.16154	.000	2.0726	3.0274
5	1	.05000	.16154	1.000	-.4274	.5274
	2	.16667	.16154	1.000	-.3107	.6441
	3	.25333	.16154	1.000	-.2241	.7307
	4	-2.60000*	.16154	.000	-3.0774	-2.1226
	6	-.05000	.16154	1.000	-.5274	.4274
6	1	.10000	.16154	1.000	-.3774	.5774
	2	.21667	.16154	1.000	-.2607	.6941
	3	.30333	.16154	.919	-.1741	.7807
	4	-2.55000*	.16154	.000	-3.0274	-2.0726
	5	.05000	.16154	1.000	-.4274	.5274

*. The mean difference is significant at the 0.05 level.

The above table 4 revealed that there was a significant difference in list four from other lists; therefore list number four was eliminated. List five and six was renumbered as list four and five.

Finally 5 Bengali sentence identification test list was standardized and the entire list contained all the phoneme of Bengali language. The mean normative performance of Bengali sentence identification test was measured at global SNR (-5dB) that bare a percentage score of 51.70%.

Test-retest reliability

Test-retest reliability of the developed sentence material was measured for all the five lists. For that, the test material was applied to the same 60 individual with one month interval from the first time of administration. Pearson correlation coefficient of their performance was assessed for each list.

Table 5: Pearson correlation coefficient of Pre and post sentence identification score at -5 dB SNR for sentence list one**.**

		List1Pre	List1Post
List1Pre	Pearson Correlation	1	.716**
	Sig. (2-tailed)		.000
	N	60	60
List1Post	Pearson Correlation	.716**	1

	Sig. (2-tailed)	.000	
	N	60	60

Table 5 represents the Pearson correlation coefficient of Sentence List one at 95% confidence interval. Table shows that there was no significant difference in the pre and post sentence identification test scores for list one. Pearson correlation coefficient was 0.71 that indicates the presence of high correlation between pre and post sentence identification test score.

Table 6: Pearson correlation coefficient of Pre and post sentence identification score at -5 dB SNR for sentence list two.

		List2Pre	List2Post
List2Pre	Pearson Correlation	1	.659**
	Sig. (2-tailed)		.000
	N	60	60
List2Post	Pearson Correlation	.659**	1
	Sig. (2-tailed)	.000	
	N	60	60

Table 6 represents the Pearson correlation coefficient of Sentence List Two at 95% confidence interval. No significant difference was observed in the pre and post sentence identification test scores for list Two. Pearson correlation coefficient was 0.65 that indicates the presence of high correlation between pre and post sentence identification test score.

Table 7: Pearson correlation coefficient of Pre and post sentence identification score at -5 dB SNR for sentence list three.

		List3Pre	List3Post
List3Pre	Pearson Correlation	1	.779**
	Sig. (2-tailed)		.000
	N	60	60
List3Post	Pearson Correlation	.779**	1
	Sig. (2-tailed)	.000	
	N	60	60

Table 7 represents the Pearson correlation coefficient of Sentence List Three at 95% confidence interval. Table shows that there was no significant difference in the pre and post sentence identification test scores for list Three. Pearson correlation coefficient was 0.77 that indicates the presence of high correlation between pre and post sentence identification test score.

Table 8: Pearson correlation coefficient of Pre and post sentence identification score at -5 dB SNR for sentence list four.

		List4Pre	List4Post
List4Pre	Pearson Correlation	1	.821**
	Sig. (2-tailed)		.000
	N	60	60
List4Post	Pearson Correlation	.821**	1
	Sig. (2-tailed)	.000	
	N	60	60

Table 8 represents the Pearson correlation coefficient of Sentence List Four at 95% confidence interval. Table shows that there was no significant difference in the pre and post sentence identification test scores for list one. Pearson correlation coefficient was 0.82 that indicates the presence of high correlation between pre and post sentence identification test score.

Table 9: Pearson correlation coefficient of Pre and post sentence identification score at -5 dB SNR for sentence list five.

		List5Pre	List5Post
List5Pre	Pearson Correlation	1	.786**
	Sig. (2-tailed)		.000
	N	60	60
List5Post	Pearson Correlation	.786**	1
	Sig. (2-tailed)	.000	
	N	60	60

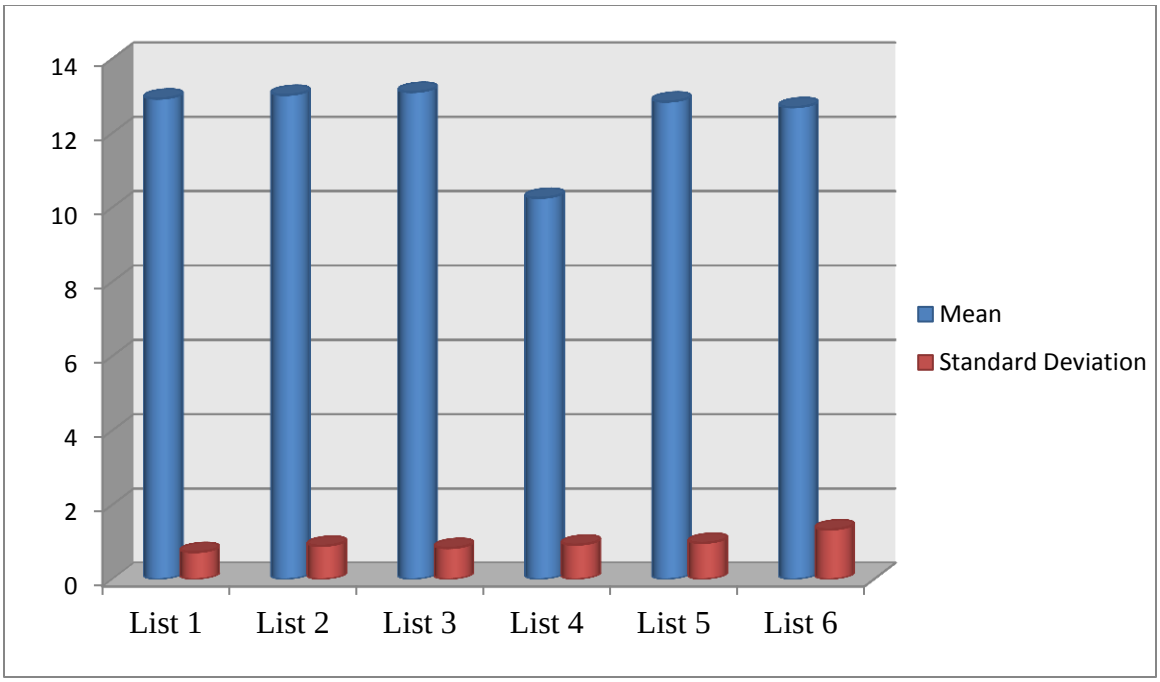
Table 9 represents the Pearson correlation coefficient of Sentence List Five at 95% confidence interval. Table shows that there was no significant difference in the pre and post sentence identification test scores for list five. Pearson correlation coefficient was 0.78 that indicates the presence of high correlation between pre and post sentence identification test score.

Test retest reliability measurement using Pearson correlation coefficient method was done with SPSS software version 16.0. No significant difference was found across the list which was suggestive of high intra list correlation.

5. DISCUSSION

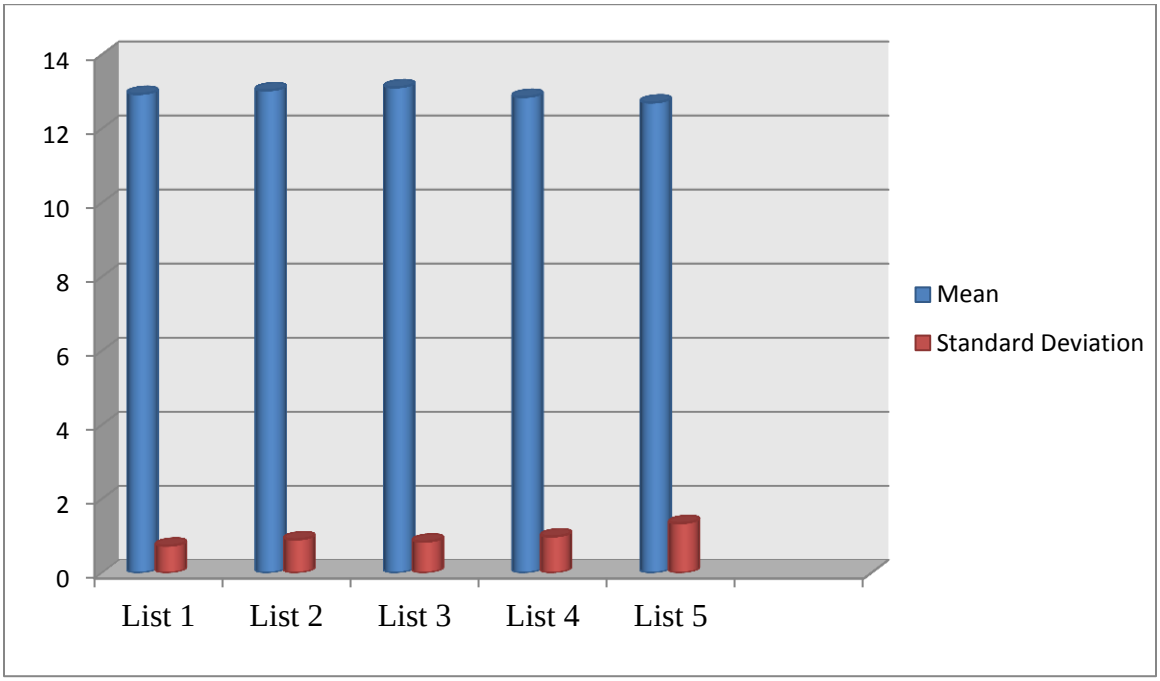
Speech perception ability is highly influenced by the language. Various studies used different numbers of words to develop sentence material in different language. Wagener et al. (2003) developed the Danish speech material named as DANTALE II. They used five words per sentence for development of the test

material In an Indian native language study done by Hota, Dutta, Chatterjee, and Sinha (2016) used five to seven words to develop sentence material in odia language. Geetha, Kumar, Manjula, and Pavan (2014) developed a Kannada sentence test material using four to six words in a sentence. One more important study in Hindi language was done by Jain, Narne, Singh, Kumar, and Mekhala (2014). They used three to seven words per sentence to develop Hindi sentence corpus. Different studies used different word range and stringently followed it. Specification of word range for each sentence was done only to maintain the homogeneity of the sentences across list (Hota, Dutta, Chatterjee, and Sinha; 2016). . For this study, 8-9 syllables in a sentence were used. Previous researchers like Vaillancourt et al. (2005) adapted the hearing in noise test procedure to develop a sentence test for Canadian francophone population. They used 5 to 7 syllable for a sentence. They suggested that lengths of the sentences should be kept short and equal to reduce the memory effect. Global SNR can be defined as the SNR level at which sentence identification score revealed approximately 50% (Kollmeier and Wesselkamp, 1997). The SNR level at which sentence identification scores came approximately 50% also defined as SNR 50 by Jain, Narne, Singh, Kumar, and Mekhala (2014). Speech perception ability deteriorates significantly in presence of background noise. Hence it is important to measure the permissible noise level at which at least 50% of the speech can be identified correctly. Sentence equivalency was assessed in order to keep all the sentences equal in difficulty. This result was comparable to the result of Kollmiere and Wesselkemp (1997). Their result revealed 20% sentence identification score at -8 dB SNR, 50% sentence identification score at -6.2 dB SNR and 80% sentence identification score at -4 dB SNR. This small discrepancy may have resulted due to the derivation of weighting factors to the sentences to reduce the in-homogeneity. A similar result was also reported by Geetha, Kumar, Manjula, and Pavan (2014) who achieve 75% score at –3 dB SNR, 50% score at –5 dB SNR, and 30% score at –7 dB SNR. Some similarities were observed in both the studies as it was concerned with the native Indian language. Standardization of the test material was done to check the homogeneity of test material and to rule out a normative value. Total 69 sentences were distributed in 6 lists, which consist of 10 sentences in each list. Remaining 9 sentences were used for familiarization of the test. This sentence list was presented to 60 normal hearing participants and their mean score of correctly identification of key words were measured. ANOVA: single factor test was done to observe the presence of performance variation across lists. Results showed that performance across list was equivalent, only list four showed different score from the other lists. Therefore, list four was eliminated.



Mean score and SD of sentence identification test for all six lists.

After eliminating the list number four, total five lists were finalised as Bengali sentence identification test material.



Mean score and SD of Bengali sentence identification test for final five lists.

Normative value of mean sentence identification score was obtained 51.70% at -5 dB SNR.
Obtained standard deviation of this study was comparable to the results of Sbompato et al. (2014). Their study revealed a standard deviation of 1.55 in noise left condition, 1.63 in noise right condition, 0.80 for

compound noise condition and 0.89 at noise front condition. This result was dependent upon the SNR level at different noise condition. In contrast, Kollmeier and Wesselkemp (1997) found higher mean and standard deviation with German sentences. These variations of mean and standard deviation suggest that this test is highly homogenous in nature and it can be used in a wide range for clinical practise. Similar study was done by Wagener (1973) in which twenty five sentence lists were developed with ten sentences in each list. Further the sentence lists were optimized and combined it into two set of list. First set consists of five list of thirty sentences and second set consist of five list of twenty sentences and were presented at ten different SNR (-18 to 0 dB with two dB increment). It was found -8.38 dB mean SNR with standard deviation of 0.16 dB across list and -8.43 dB mean SNR with standard deviation of 0.95 across subjects.

Test retest reliability for any test is to monitor the consistency of the test result over time. Plomp and Mimpen (1979) suggested that a test can be said as reliable if a little difference found between the test and retest score. In this study the test was administered to the same subjects after one month and their sentence identification score was calculated. Pearson corelation coefficient was performed for each list to check the reliability of the test list. Table 9 demonstrates the Pearson correlation coefficient of 0.78 for list five which suggested a high correlation between pre and post sentence identification scores.

6. CONCLUSION

The study was conducted in a well established audiological set up. Methods incorporated subjective naturalness and predictability measurement of sentences, objective measurements of audiological tests, assessment of test retest reliability of the sentence list. All the test materials were presented only once to each individual to reduce the fatigability and learning effect. Developed sentence lists have a wide range of implications.

1. Sentence lists can be used to assess the speech perception of Bengali native speakers.
2. It can be used as a test battery for routine hearing assessment.
3. Hearing aid performance also can be assessed by comparing the pre hearing aid fitting result with the post hearing aid fitting scores.
4. Comparison of hearing aid performance will also be possible with Bengali sentence identification test.
5. Bengali sentence identification test can be used to monitor the progress of rehabilitation.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Communicative Form vs. Literary Form: An intervention inspired by Nigel Fabb

Rimi Ghosh Dastidar
P.R Thakur Govt. College, West Bengal, India

ARTICLE INFO

Article history:
Received 27/07/2020
Accepted 06/08/2020

Keywords:
Communicative Form,
Literary Form,
Prosody,
Relevance Theory

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Prosody is conventionally considered to be the inevitable part of poetic pieces in Literature and is set to be analysed on the basis of statistical measurement through standard method. This so-called regular formula often deviates depending on the contextual variation of the particular literary content. This paper will focus on the meeting point between literary form and communicative form in respect of poetic elements and other than poetic elements in respect of literature. It tries to explain how human mind reacts to different literary forms by the presence of prosody and how it relates every day habitation of language. We try to find out this variance of prosodic occurrence through the point of view of Nigel Fabb.

1. Introduction:

Nigel Fabb (2004) introduces us with the term Literary Form which actually suggests that "A text has Literary Form if certain statements are true of the text." If we consider the ground of poetry, we find variety of forms of texts with different names. In English, a poem may be tagged as sonnet if and only if it has 14 lines with 8-6 couplet division and allows iambic pentameter to measure its metrical statistics. In Bengali a poem becomes a rhyme if it shows end rhythmic pattern with lucid flow of prosody and employs syllabic pattern as metrical tool. A poem turns to be epic if it is set in a long narrative style and moraic/ mixed moraic metrical pattern. It is not only the case of poetic pieces; whatever genre we find in the ground of literature, each genre maintains certain principle to some extent to be a distinctive profile. This profile is marked as a

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Rimi Ghosh Dastidar
Email: rimigd@gmail.com

specific literary form. Therefore each and every text carries some specific features which help them to be identified as a definite name. Fabb (2004) also points out that there exists another form, namely communicative form which acts like a bridge between a literary piece and a performer (Reader/ Audience/ Speaker).How the reader conceptualizes a text is determined by this communicative form or vice versa. When literary form meets communicative form, prosody becomes the prior tool to materialize the text. This prosodic endeavour corresponds not only with poetic elements; it makes an immediate connection with our regular conversation. We often ignore the implied existence of prosody within regular conversation, whereas conversation also needs prosodic scansion to be explained in its particular context of occurrence.

As we mentioned before, a sonnet in English is usually formed with 14 lines with 8-6 division and employs iambic pentameter as metrics. Iambus is considered as the foot in English prosodic scansion where unstressed syllable occurs first and stressed syllable appears as the immediate follower of the preceding one. Pentameter marks the measurement of five metre. A typical English sonnet feeds this mechanism with precision. According to Fabb (2004) it is true that this machinery works for sonnet but what happens if a sonnet applies such a content which requires 6-8 division to be explored properly rather than 8-6 division? What should we supposed do in this type of case? Should we prefer conventional form as decided by grammaticality or should we consider the flexibility as desired by the content?

2. Implied, Explicit and Generated Metrical Form:

Fabb (2004) shows that three forms are there in the structure of a poem- 1) Implied form 2) Explicit Form 3) Generated metrical form. Implied form refers to that definite form which clearly exposes a poem in a specific genre. Whatever it may be! Either sonnet or rhyme! 'Degree of Intensity' is the crucial factor in case of implied form. 'Degree of intensity' considers the constraint of parameters of a poem to categorize it under a specific term. Explicit form is the surface structure of a poem. Fabb (2004) labels this form as performance, i.e, how the poem is exposed to the reader? Is it either in printed version or in spoken form? A text is realized differently depending on the nature of performance. Let us consider example (1)

(1)

tomar bondhu ke? dirghoSSaS?
amar o tai
amar sunnota gOnonahin
tomar o tai?

(kOthopokOthon, Purnendu Patri)

[Who is your friend? Sigh?
I have too
My emptiness is coutless
Is that your too?]

(Conversation, Purnendu Patri)

When reader finds this poem in written form, like you, she realizes this text in a particular way because the words are abstract here. When these words are recited, comprehension level may differ depending on the speaker's emotion and listener's reaction because the words remain living in our spoken appearance. If you find the word dirghoSSaS [sigh] in print, you read it as a mere word. When you hear the word from someone, you may perceive the contextual implication of the word.

Generated metrical form calculates the statistical measurement of metrics in prosody within a poem. Some conditions and conventions build up this generated metrical form. These three forms overall generate prosodic wave within a poem. Therefore prosody is not an isolated entity. Totality of a poem, with which, a poem makes its appearance before the reader, determines its prosodic behavior.

3. Literary Form and Communicative Form in Poem:

Let us find example (2) to find out the anti-regular break of sonnet in Bangla, as Fabb (2004) explains it in English.

(2)

bhije hoye aSe meghe e dupur- cil Eka nodiTir paSe	a
jarul gacher Dale boSe boSe ceYe thake oparer dike	b
payra giyeche ure cobutOre, khope tar SoSa lota Tike	b
chere gEche moumachi –kalo megh jomiache magher akaSe	a
mOra projapotiTir pakhar nOrom renu phele diYe ghaSe	a
piMpRera cole jay dui dOnDo amgache Salikhe Salikhe	b
jhuTopuTi kolahOl bOukOthakOw ar raNaTike	b
Dake nako holud pakhna tar kono jEno kaMThal pOlaSe	a
haraYeche, bowo to uThane nai pore ache Ekkhana DheMki	c
dhan ke kuTibe bOlo-kOtodin Se to ar kate nako dhan	d
rodeo sukute Se je aSe nako chul tar kOre nako snan	d
e pukure bhaMRare dhaner bij kOlaye giYeche tar dekhi	c
tobuo Se aSe nako, ajo e dupure eSe khoi bhajibe ki	c
he chil, Sonali cil, raNa rajkOnna ar pabe na ki pran.	d

[bhije hoYe aSe meghe e dupur, Jibanananda Das]

Noon becomes drenched with cloud- a kite, all alone, at the bed of the river
Sitting on the branch of Jarul Tree Looks on to the side other
Pigeons have flew to sky; in the honeycomb lies the creeper
Bees have left- dark cloud mounds in the blue of winter
Leaving the soft dust of the wings of dead butterfly
The ants move on- Salikhs are on the Mango Tree for a while

Flittering and arguing with each other- bou-kOtha kow
Stops calling the new bride, her yellow wings in some Jackfruit and Palash

Have been lost. Bride is not in the yard. Only the husking pedal is there.
Tell me who will reap the rice? As long time she didn't reap rice.
She does not come to dry her hair in the Sun. Takes no bath
In this pond and coffer, the seeds of her paddy sprouted and piled up
Still she does not come! Will she come midday to fry the parched rice?
Oh! Kite, golden kite, will the brightened princess anymore revive?

[The Noon becoming drenched in cloud, Jibanananda Das]

The poem is of 14 verses with 8-6 division and we find a list of end-rhyme abba (/Se/-/Ke/-/Ke/-/Se/(, abba (/Se/-/Ke/-/K(h)e/-/Se/), cddc (/ki/-/an/-/an/-/k(h)i/), cd (/ki/-/an/). Since eighth verse, same end rhyme occurs whereas it changes from ninth verse. According to the structural phenomenon, the poem must be a sonnet as we discussed in implied form. 'Degree of intensity' stimulates this form to be the sonnet. But the fact is, if we look at the content of the sonnet, the lingering effect of last two words (kaNThal, 'jackfriut' and pOlaS, 'flower') of eighth verse cannot be completed until and unless the word hOlud pakhna (yellow wing) of eighth verse relates with the word haraYeche of ninth verse at a single flow. If we follow end rhyme structure, we have to stop at eighth verse, whereas if we look at content, we must break the sonnet into 4-10 division to maintain a logical weaving within the content
In context of explicit form, we may say that whenever the sonnet appears in written version withsome abstract word, we don't find any deformity in the structure and accept the 8-6 division. But when we read the sonnet, we subconsciously fuse 8th and 9th verse in flow of content breaking the conventional division. We utter at a single move, *holud pakhna tar kono jEno kaNThal pOlaSe haraYeche.*

To say about the generated metrical form, we find that this poem follows moraic patter in its metrical measurement where open syllable is counted as one metre and closed syllable is counted as two metre.

1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	1
bhi	je	ho	Ye	a	Se	me	ghe	e	du	pur	cil	E	ka	no	di	Tir	pa	Se
1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
ja	rul	ga	cher	da	le	bo	Se	bo	Se	ce	Ye	tha	ke	o	pa	rer	di	ke

4. Literary Form and Communicative Form In Dialogue from Literature:

Let us come to the second part of our discussion. Now we look at another genre of literature, namely Dialogue and in this context we will consider *am aMtir bhepu* by Bhibhutibhusan Bandyopadhyay. *am aMtir bhepu* is purely a canvas of Bengal village where Apu and Durga,

two child characters play the role of siblings (brother and sister). We will focus on their conversation and find out how prosody works in the literary conversation. We have already discussed about literary form. Now let us have a discussion communicative form. Communicative form is not completely distant from literary form. Rather communicative form can be generated through literary form. Some kind of literary forms are there where set of laws is strictly imposed and chances of deviation are hardly found. On the other hand, some literary forms consider content as its primary tool and decrease the presence of constraints. Consequently, we find deviated form of that literary form. According to Fabb (2004) "...Literary form of this kind has no objective existence in the text". How the reader communicates with the text would be the best way for the text to be recognized. This mechanism is communicative form, i.e, communication between reader and literature. Fabb (2004) says that communicative form can be analyzed through linguistic pragmatics. Here we would not elaborate the topic on linguistic pragmatics; rather, we observe the role of prosody in communicative form. In *am aMTir bhepu* Apu and Durga converse in such a way, where writer associates expression with the characters before and after their speech. Therefore, dynamics of prosody is reflected not only through the chain of the words but also through the adherence of the speech. And reader can communicate with the sentiment of the characters. Let us consider instance (3)

(3)

When Durga arrives home after spending a long hours outside, she calls her brother with a wary tone – 'Apu, o Apu'. Apu attentively responds with low voice- ki re didi ?

Durga: ma ghaT theke phere ni to?
Apu: uhu
Durga: ekTu tel ar nun niye aSte pariS? mer kuSir jObab.
Apu: kotha theke peli re didi?
Durga: poTlider bagane SiMdur kouTor tOlaY poRechilo. an diki ekTu nun art el
Apu: teler bhaR chule ma marbe je! Amar kapoR je baSi
Durga: tui ja na Siggir kore, aSte Ekhon Dher deri.. khar kacte giYeche- Siggir ja
Apu: narkeler mala Ta amaY de. Ote Dhele niYe aSbo, tui khiRki dore giYe dEkh ma aSce kina
Durga: tel-Tel jEno mejhete DhaliS ne, Sabdhane nibi, noile ma Ter pabe, tui to EkTa haba chele

Durga: Mother did not yet return from ghat?
Apu: No
Durga: Can you ring some oil and salt. Small green mangos are here.
Apu: Where have you found these?
Durga: Those were fallen under box of vermillion in the garden of Patli. Bring some salt and oil.
Apu: Mother will beat me if I touch the jug of oil. My clothes are not washed still.
Durga: Go fast. Mother will take time to come. She has gone to wash clothes. Go fast.

Apu: Give me the shell of the coconut. I will take oil in that. You may go and watch if mother comes or not.
Durga: Be aware! Don't pour single drop of oil on floor. Take it carefully otherwise mother will doubt. You are an idiot guy!

The text is recognized as a dialogue, type of a literary form because it incorporates all the required features for being categorized as dialogue. Fabb (2004) says, "Type of literary form holds a text as the content of an inference with certain degree of strength." But here we view the internal structure of the dialogue. Whenever we speak, we never follow the pre-determined metrical system because nature of spoken language is spontaneous. Prosody must exploit its whimsical nature. Poem has a definite form with a layout of words. Regular conversation is built up with spontaneous words required contextually. Let us explain the dialogue. When durga asks whether her mother has returned from *Ghat* and gets a negative reply from Apu, she subsequently plans her work. When Apu becomes alert, apprehending the scolding of mother, Durga immediately assures him that mother will take time to return and Apu springs up and runs to carry out his sister's order.

Can we consider the dialogue only as a literary form produced with some words? No. It's not just the abstract construction of words. Each and every word is floated with a prosodic sense. At the very beginning, when Durga calls her brother, we find a low tempo with a hastening mode. Again, we find Durga to say 'an diki ekTu nun ar tel' or 'tui ja na Siggir kore', these words move with a speed because Durga knows that she has to do it rapidly lest her mother comes and they will be in the trap. Reader can assimilate herself with the dialogue, not only through the meaning of the words, but through the prosodic diversity of the words. Fabb (2004) wishes to observe this area as implicature. "An implicature is an implication which is intentionally communicated". There must be some stimulating feature within a dialogue which helps the reader to communicate her cognitive sense with the context and meaning of the dialogue. Fabb (2004) also connects the Relevance Theory (Sperber and Wilson 1986, 1995(with implicature. According to him, the production of implicature is guided by the principle of relevance. Dialogue lacks its spontaneous mode if no symmetrical framework exists within the movement of context, word and prosody. When Durga makes her brother alert by saying, "tel-Tel jEno mejhete DhaliS ne, Sabdhane nibi, noile ma Ter pabe, tui to EkTa haba chele", we notice fearful tone in the voice of Durga and at the same time, we find an affectionate tone which elevates the degree of bonding between the siblings.

Fabb (2004) refers to Fodor (1975) in context of Relevance Theory. Fodor (1975) says about Relevance Theory that Thoughts are propositions, instantiated as sentences in a language of thought. Fodor classifies perception in two forms: 1) Modularized 2) General. Former one is restricted and latter one is flexible. The second form allows the reader to communicate deeply with the literary form and we easily extract the contextual variety of emotion of the characters.

Let us consider another dialogue of Apu and Durga where we find an absolute different shade of feeling:

(4)
Durga: Son Apu, EkTa kOtha Son
Apu: ki re didi?
Durga: Sere uThle amaY Ekdin tui railgaRi dEkhabi?
Apu: dEkhabo Ekhon, tui Sere uThle babake bole amra SOb Ekdin gONga naite jabo railgaRi kore.

Durga: Apu , listen!
Apu: What happens, Sister?
Durga: Will you take me to see Train when I will be cured.
Apu: I will take you Sister. I will speak to father and we will all travel by train for a bath in Ganga Ghat

When the dialogue occurs, Durga is severely ill and talks with her brother in a feeble voice. Her address to Apu completely differs from the previous dialogue where her every word reflects charming and frolic attitude. Example (4) presents a matured tone in Durga's voice and Apu's words become steady and supportive in prosodic nature, as if, circumstance makes both the sibling calm and quiet. And all these variations are mirrored through the prosodic wave of dialogue.

5. Literary Form and Communicative Form In Song:

Now we will look into another literary form, that is, Song to see how this genre works as communicative form to the audience. Let us take (5) as our instance in this respect.

(5)
prothomoto ami tomake cai
ditiyoto ami tomake cai
triyoto ami tomake cai
SeS porjonto tomake cai
nijhum Ondhokare tomake cai
ratbhor hole ami tomake cai
SOkaler koiSore tomake cai
Sondher ObokaSe tomake cai

…

odhikar bujhe neoa prokhor dabite
Sararat jege aka lOraku chobite
chipchipe kobitar chOnde bhasaY
godder juktite baMchar aSaY
srenihin SOmajer ciro baSonaY
dinbOdoler khide bhOra cetonay
didhadOnder din ghocar SOpne
Sammobader gan ghume jagorone

biplObe bikkhobhe tomake cai
bhiSon OSombhObe tomake cai

[tomake cai, Kabir Suman]

First of all, I expect you
Secondly I expect you
Thirdly I expect you
Till the end I expect you
…
Grasping Strong understanding to achieve right
Drawing combative picture during sleepless night
In the rhythmic language of constricted verses
Hope to live in the reasons of prose
With the eternal desire for classless society
Longing for change of epoch in full sensibility
In hope to wiping out the days of dubiety
Conscious or not, lipping the song of equality
I expect you in Tumult and Mutiny
I expect you in extreme adversity

[I expect you, Kabir Suman]

Now let us consider the difference between two stanzas and how audience relate the difference. Verses of First stanza purely show romanticism for the beloved. The words as well as the prosodic wave expose a soft approach of the singer to the world. While we move to the next stanza, the words register revolutionary attitude and the prosody also alters its direction on an uprising mode. Therefore the audience also communicate themselves with the dynamic range within a single song. This is the communicative form which connects the reader/ audience / perceiver with the literary form.

6. Conclusion:

Therefore we conclude here with the inference that a text requires both structural pattern to be classified as an entity and content to be exposed to the readers. Structuralism and esthetics corresponds to each other and formulate a text. Until a literary form remains abstract in written format, it communicates with the reader in a structural configuration. When the text is uttered, it connects the reader in a live set-up. Revolutionary Theatre movement reinforces that Form is worthless if it is not the form of its content (1982). So the motion of text must be discursive so that the reader seizes the ambience of the text intensely.

Reference:

Bandyopadhyay ,Bibhutibhusan .2018 .Am aMTir bhepu. Sishu Sahitya Sangsad। Kolkata

Das ,Jibanananda.2017 .SreSTho kobita. Dey'z publishing, Kolkata

Kabir Suman.1992. tomake cai. Tomake Chai Album. HMV.Kolkata

Potri ,Purnendu.2007.kOthopokOthon .Dey'z publishing, kolkata

Fabb. Nigel. 2004. Language and Literary Sturcture. The Linguistic Analysis of Form in erse and Narrative. Cambridge University Press. UK
Fodor, J .1975, The language and Thought . Tomas Y. Crowell. New York.
Fodor, J .1983, The Modularity of Mind. MIT Press. Cambridge.
Sperber, Dan and Wilson.1986. 2nd Edition 1995. Relevance: Communication and Cognition. Blackwell. Oxford.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Improvisation in Indian Classical Music: Probing With M-B and B-E Distributions

Souparno Roy[1], *Archi Banerjee*[2,1] *and Shankha Sanya1*[1]
[1]Jadavpur University, India; [2]IIT Kharagpur, India

ARTICLE INFO

Article history:
Received 14/05/2020
Accepted 13/11/2020

Keywords:
Maxwell-Boltzmann distribution,
Bose-Einstein distribution,
Indian classical music,
Raga,
improvisation,
temperature

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Indian classical music (ICM) is considered to be one of the oldest and mostcreative art forms existing in this world and *Raga*, in spirit, is the structural unit that binds together the vast expansion of this music genre. A musician while performing expresses the *Raga* according to his mood and environment surrounding him. Thus the performances, even with the same Raga, have some subtle differences with each other. These differences are generally called "improvisation". In this study, we intend to quantify such improvisations using measurable parameters inspired by statistical tools used in Physics. To study them quantifiably, we introduce methods based on well-known concepts of Statistical Physics (especially thermodynamics), namely Maxwell-Boltzmann (MB) statistics and Bose-Einstein (BE) distribution. In this present study, these distributions have been applied to find new parameters (equivalent to 'temperature' in physical systems) to identify different features of improvisation in different Hindustani classical music performances of the same Raga by the same artist. Music clips chosen were 6 different renditions of the same *Vilambit bandish* of the same Raga sung by legendary classical music maestro Kumar Gandharva on 6 different occasions. The resulting analysis gives a number of parameters (they come from the analogy between the rank-frequency distribution and the respective statistical distribution) that help in identification and categorization of the improvisational changes in the chosen 6 renditions and thus parameters such as individual improvisation pattern which were previously considered as abstract and unexplainable to naive listeners (individuals without musical training in ICM) can now be analysed from a quantitative approach. The methods studied here are novel in the music research field and can prove to be useful in the fields of music and speech as quantifying parameters for artist style and Raga identification.

1. Introduction

Indian classical music (ICM) is considered to be one of the oldest and most creative art forms existing in this world and *Raga*, in spirit, is the structural unit that binds together the vast expansion of this music genre. The word *Raga* is derived from the Sanskrit word "*Ranj*" which literally means to delight and gratify [1][2]. Although there are a number of definitions attributed to a *Raga*, it is basically a diverse tonal module. The listener has to listen to several pieces of the same *Raga* in order to recognize it. Each

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Souparno Roy
Email: thesouparnoroy@gmail.com

Raga has a well-defined structure consisting of a series of four/five or more musical notes upon which its entire melodic expression is constructed [3]. However, the way of approaching and rendering the notes in musical phrases and the mood they convey more often defines a *Raga* than the notes themselves. Basically, the performers of ICM visualize every *Raga* as a living existence. A musician while performing expresses the *Raga* according to his mood and environment surrounding him. During a performance, a singer steadily loosens the stranglehold of the rules of music in a subtle way. He does not disregard them, but merely interprets them in a new way, which is the beauty of Indian classical music, where *Raga* and its grammar are only means and not ends. Every performer of this genre is essentially a composer as well as an artist. Thus, even if an artist sings or plays the same *Raga* and same *Bandish* twice, there are always certain differences from one rendition to another. These differences are generally called "improvisation". Musical improvisation is generally defined as the creative activity of immediate ("in the moment") musical composition, which combines performance with communication of emotions and instrumental/vocal techniques [4]. So, the term "improvisation" often resembles with the concept of extemporisation in case of ICM. Improvisation is a common form of musical practice across cultures, and yet remains scarcely studied or understood from a scientific musical analysis point of view. In ICM, when an artist performs a certain *Raga*, keeping the structural framework of the raga intact, the existing phrases are stretched or compressed, and the same may happen to motives from the phrases; further motives may be prefixed, infixed and suffixed. These characteristics throw a vast number of musical information towards the listener and analysts alike. When looked at as a whole, this congregation appears too complex. But, similar to every other information humanly perceivable, it also is made up of repetitive pattern or sequences of some common basic elements. One central problem in the analysis of these sequences is how to effectively categorize their information content based on the common elements found in their origins. Variations of musical compositions - be it in rendition styles or in emotions it conveys- can only be recognized by trained and experienced listeners. This kind of categorization is mostly non-quantifiable. Hence, to categorize such musical information, one needs to address the specific problems of identifying the sequential patterns and quantitatively applying this knowledge in subsequent comparison. In this work, following the arguments above, we have attempted to quantify such abstractness using measurable parameters. For that, we introduce methods based on well-known concepts used in Statistical Physics (especially thermodynamics). This study looks to investigate the concept of improvisation in ICM from a scientific and universal point of view using two of the most fundamental distributions of statistical physics, namely Maxwell-Boltzmann (MB) statistics and Bose-Einstein (BE) distribution.

In the last few decades, the usage of Statistical tools in social sciences, linguistics and structural biology has gained attention, since they usually deal with large data sets [5]-[8]. The basis of the application of the statistical methods in the structured, high dimensional data is a brilliant empirical law, primarily used in literature research, known as Zipf's Law, formulated by linguist George Zipf in 1949 [9]. It states that if we we assign the rank j = 1 to the most frequent word of a language, j = 2 to the second one, etc., then the frequency of occurrence f(j) of a given word varies with its rank j as:

$$f(j) \sim 1 / j^{\alpha} \tag{1}$$

; where α is an exponent which is to be determined from the rank vs frequency distribution.

Zipf's law is strikingly remarkable because it can be applied to diverse systems, including economics, linguistics, even urbanization models [10]-[13]. It is also hard to overlook its similarity with

statistical distributions concerning energy of a system of particles in equilibrium. According to statistical mechanics, when a system of particles is in equilibrium at constant temperature T, then it can be found in one of N states permissible. The probability p_i that it is found at a given state i with energy E_i is:

$$p_i \sim 1 / \exp(\beta E_i) \quad (2)$$

; where $\beta=1/kT$; k is the Boltzmann constant ($1.38x10^{-23}$J/K) and T is absolute temperature, the 'measure' of the interaction of the system with the environment.

The approach we follow in this study is based on the analogy between the rank-frequency distributions (using Zipf's law) of a note-duration combination of a music sample and the statistical distributions (both the Maxwell-Boltzmann distribution and the Bose-Einstein distribution in grand canonical formulation). The distribution of the occurrence of notes (combined with their durations) is characterized by a set of parameters, one of which includes the equivalent of Temperature in case of physical systems. This 'temperature' parameter is quite familiar in linguistic research [14][15] and it has been used to specify the underlying dynamics of various languages, authorship disputes, changes in complexity of vocabulary and many more. Here, we apply this statistical approach on different note-duration combinations present in the experimental samples from Indian classical music to find out whether it enables us to distinguish those different renditions according to their note distribution and movement patterns.

2. MAXWELL-BOLTZMANN AND BOSE-EINSTEIN DISTRIBUTIONS IN BRIEF

Maxwell-Boltzmann distribution

The Maxwell-Boltzmann (MB) statistics, derived in late 1800s, is used for distribution of an amount of energy between identical but distinguishable particles. In a nutshell, MB statistics predicts that for a given temperature T, for a system consisting of a huge number of non-interacting distinguishable particles the probability of a particle having larger energy decreases exponentially. The distribution function has the following form:

$$f(E_i) = 1/Ae^{Ei/kT} \quad (3a)$$

where $f(E_i)$ is the probability of a particle having energy E_i, A is the normalisation constant, E_i is the energy of the i-th state, k is the Boltzmann constant, and T is absolute temperature. By knowing Maxwell–Boltzmann distribution, we can understand and interpret the measurable macroscopic properties of materials in terms of the properties of their constituent particles and the interactions between them, i.e., we can predict observable macrostates from microstates.

Bose-Einstein distribution

Theorized in 1924/25 by Satyendranath Bose and Albert Einstein [16], Bose-Einstein (BE) statistics describes the dynamics of an ensemble of identical and indistinguishable particles occupying discrete energy states. The distribution function indicating the energy distribution looks like:

$$f(E_i) = 1/[Ae^{Ei/kT} - 1] \quad (3b)$$

where $f(E_i)$ is the probability of a particle having energy E_i, 1/A denotes the degeneracy, i.e., how many particles are having particular energy state E_i, E_i is the energy of the i-th state, k is the Boltzmann constant, and T is absolute temperature. BE distribution applies to a very particular kind of particles who have integer spin values, known as Bosons. They do not obey the Pauli's exclusion principle and hence unlimited number of particles can occupy the same energy state (The particles that obey the Pauli's exclusion principle are called Fermions). This unusual property of the BE distribution helps in applying this concept in various systems beyond the sub-atomic world.

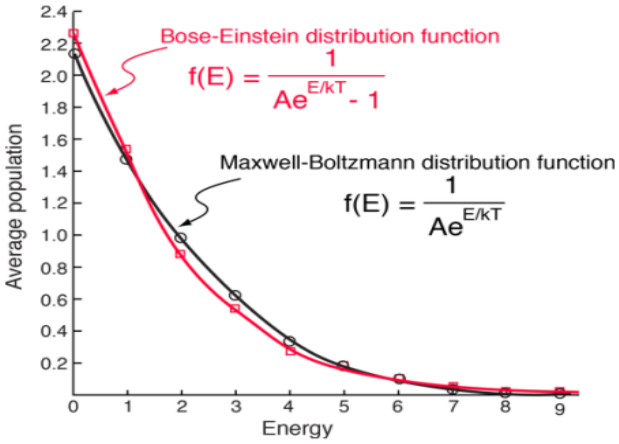


Fig. 1: Maxwell-Boltzmann and Bose-Einstein Distributions [17]

Fig. 1 depicts both the MB and BE distribution patterns.

3. APPLICATION OF MB DISTRIBUTION IN ANALYZING MUSIC

General Methodology

To analyse the Ragas, first, we need to accumulate a 'musical corpus' of used notes in the raga renditions, similar to a literary corpus (in our case it is the compilation of notes and their respective durations used in each rendition) [14]. From the pitch profile of the music sample, an experienced classical musician determines the frequency of 'sa'. Frequencies of the rest of the notes are found out with their respective frequency ratios with 'sa'. Afterwards, the existence and duration of occurrence of all the notes is indicated by analyzing the pitch profile of the music clip using Wavesurfer software. The window is taken to be 10 milliseconds each. This way we find the number of occurrences of each note and their respective durations. For example: suppose Usa_{50}, indicating the occurrence of the upper octave 'sa' for 50 milliseconds, has occurred 10 times during a piece. Similarly, Lre_{30} (lower octave 're', 30 ms duration) appears 24 times, ga_{40} (for middle octave) for 61 times etc. Then the probabilities of the occurrence of

note-duration combination are plotted along with their respective 'energies' using equation (2), taking K=T=1. This leads us to the Fig.2:

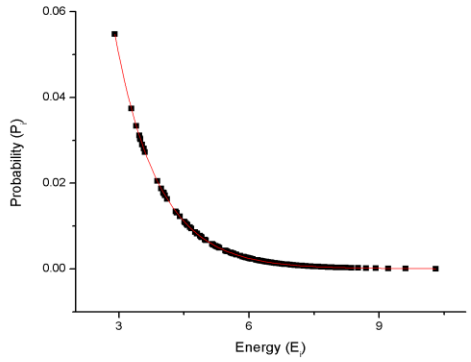


Fig. 1: Probability of occurrences of notes vs 'Energy' graph

As it's clear from Fig. 2 that the probability vs energy graph is perfectly exponential (as expected) and follows MB distribution. The temperature T of the corpus is assumed to be 1K, for comparison purposes later. The model equation we use to curve fitting is [14]:

$$p(E) = y0 + A1.\ Exp(-E/t1) \quad (4)$$

Next, we shall plot the probability (or frequency of occurrences) vs energy graph for experimental clips and derive T_{MB} in each of the cases to compare with the present 'music corpus'.

4. APPLICATION OF BE DISTRIBUTION IN ANALYZING MUSIC

Rank-Frequency distribution

To apply BE distribution, as discussed in the previous sections, first we need to prepare the rank-frequency distribution of the combination of notes and their durations. After segregating the note-duration combinations, we construct the rank-frequency distribution from these data. The component having highest number of occurrences is given rank 1; the second most frequent is given rank 2, and so on. Components with the same frequency are given a consecutive range of ranks, the ordering within which can be arbitrary.

The rank-frequency distribution of such a sample looks like Fig. 3:

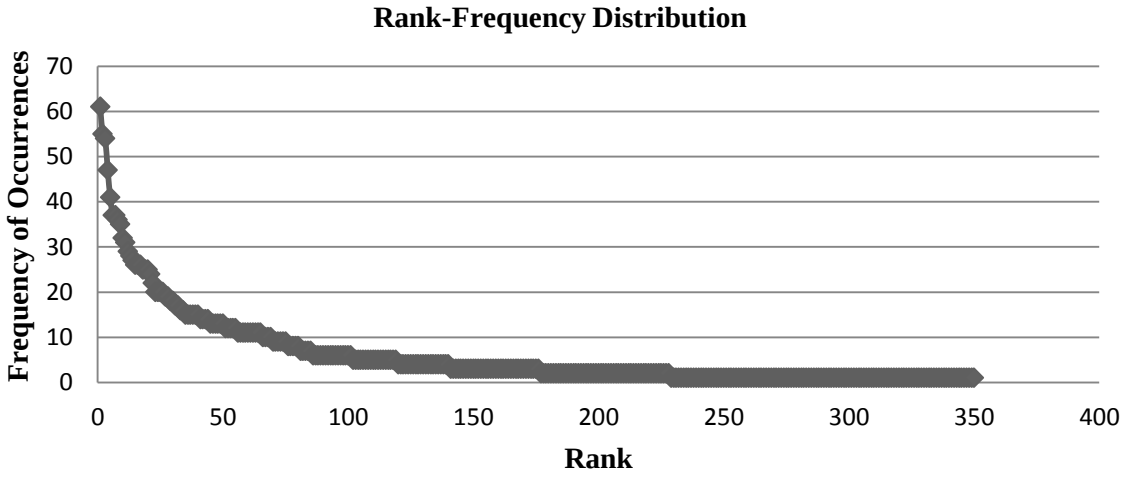


Fig. 3: Sample rank-frequency distribution for a 1 minute bandish of raga Shree

Horizontal plateaus in the domain of high ranks/low frequencies correspond to a large number of components having the same frequency. The longest plateau corresponds to frequency 1.

Physical analogy of frequency structure and B-E distribution

Following the treatment in [18], we invert the rank-frequency distribution in a relation between number of occupants N_j vs their absolute frequencies j. We identify the energy level j with the number of occurrences of note/duration combinations. Hence, the components occurring once is situated in energy level j = 1, twice occuring components sit in energy level j = 2 etc. Each of the energy levels can have any number of occupants (N_j) without any restrictions. This idea alignes in accordance to the B-E distribution, where each energy level can be occupied by any number of particles, without restricting laws like Pauli's exclusion principle. Such a plot of N_j vs j also follows the B-E distribution closely. For such a distribution, the relation between occupancy number of j-th energy level N_j and j is [19]:

$$N_j = \frac{1}{z^{-1}e^{\varepsilon_j/T} - 1} \tag{5}$$

Where, z is the fugacity, ε_j is the energy of the j-th level and T is the temperature [19]. The spectrum of ε_j is given by:

$$\varepsilon_j = (j - 1)^{\alpha} \tag{6}$$

Unity is subtracted to make sure that the lowermost energy state, j = 1, has zero energy. The main focus of the study is on the lower frequency data since the energy spectrum relationship can look different for higher energies, i.e., higher occurrency states.

Parameters to be determined

First, z is calculated from the lowest N_j value, i.e., the occupancy of the lowermost occurent state using the equation (5), putting j = 1 :

$$N_1 = 1 / (z^{-1}-1) = \frac{z}{1-z} \quad (7)$$

Also, exponent α of (6) is to be determined by fitting the plot to equation (5), along with T_{BE}.

5. RESULTS AND DISCUSSIONS

In this work, we have chosen to study six renditions of Raga *Sur Malhar* performed by Pandit Kumar Gandharva, a classical music maestro, in different times of his career. This raga belongs to Thaat *Kafi* and the characteristic notes that accompany this Raga are: Sa/high Sa, Shuddha Re (Re2), Shuddha Ma (Ma1), Pa, Shuddha Dha (Dha2), Komal Ni (Ni1), Shuddha Ni (Ni2). The duration of the sample music clips were of 4 minute, taken from the *Vilambit Bandish* part of the raga, selected by a classical music expert. *Bandish* provides the literature ingredient of the raga in traditional structured singing. Hence, in addition to keeping the raga structure intact, it provides both lyrics and melody dependent improvisations made by the artist in each of the renditions.

The M-B distribution plots:

The probability of the note-duration combinations for each raga rendition is plotted against the corpus energy of Fig. 2 and fitting a curve using eq. (4), we obtain the temperature T_{MB}. The resulting fitting plots of six music clips are given in Fig. 4:

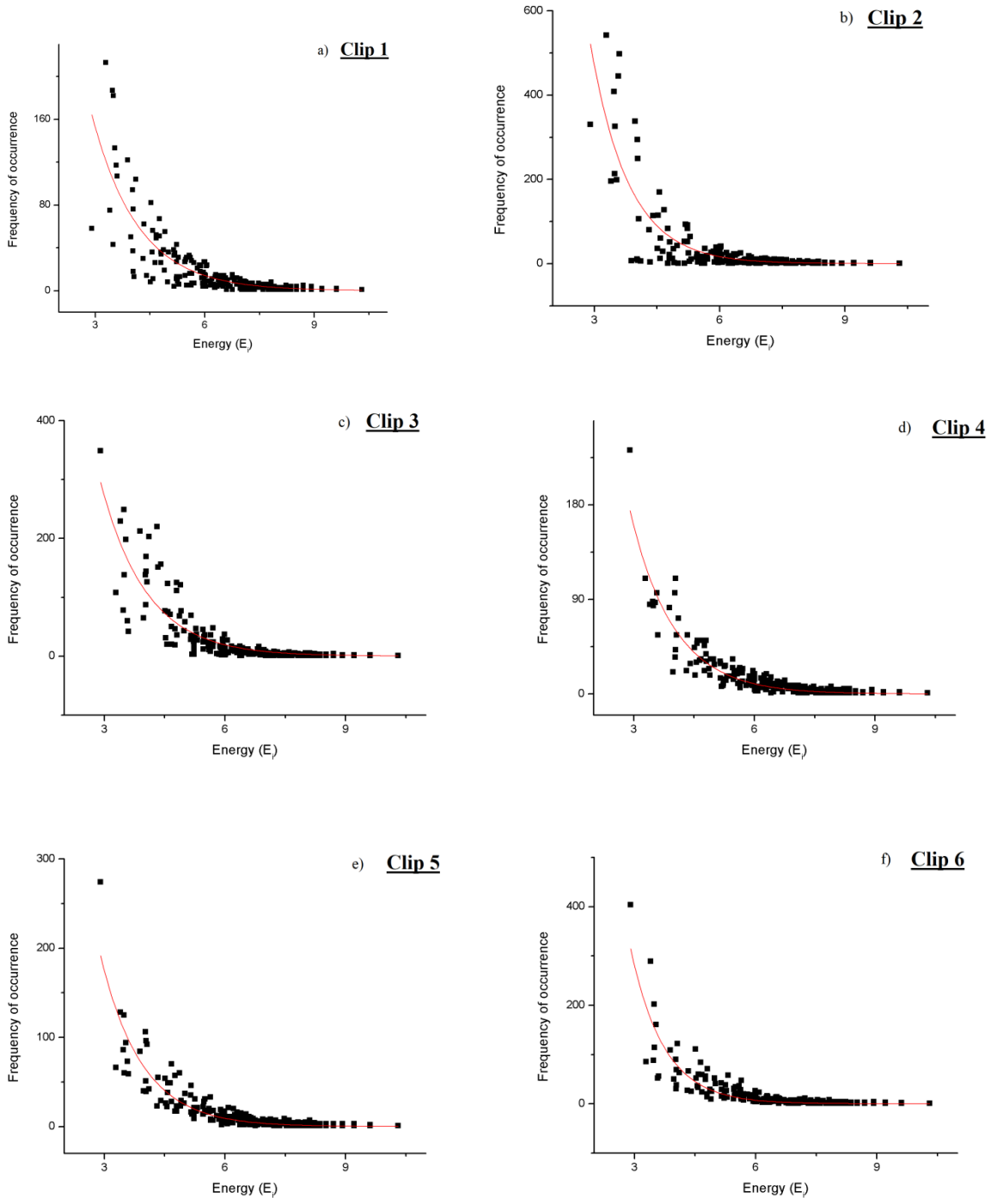


Fig. 4 (a-f): Freq. of occurrences vs Energy plots for six experiment clips

Clip No.	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5	Clip 6
R^2 value	0.71	0.75	0.78	0.87	0.83	0.79

Table 1: R^2 values of Frequency vs. Energy plots of MB distribution for six clips

The curve fitting results are given in Table 2.

The B-E distribution plots:

The N_j vs j plots for each of the six clips are given in Fig. 5:

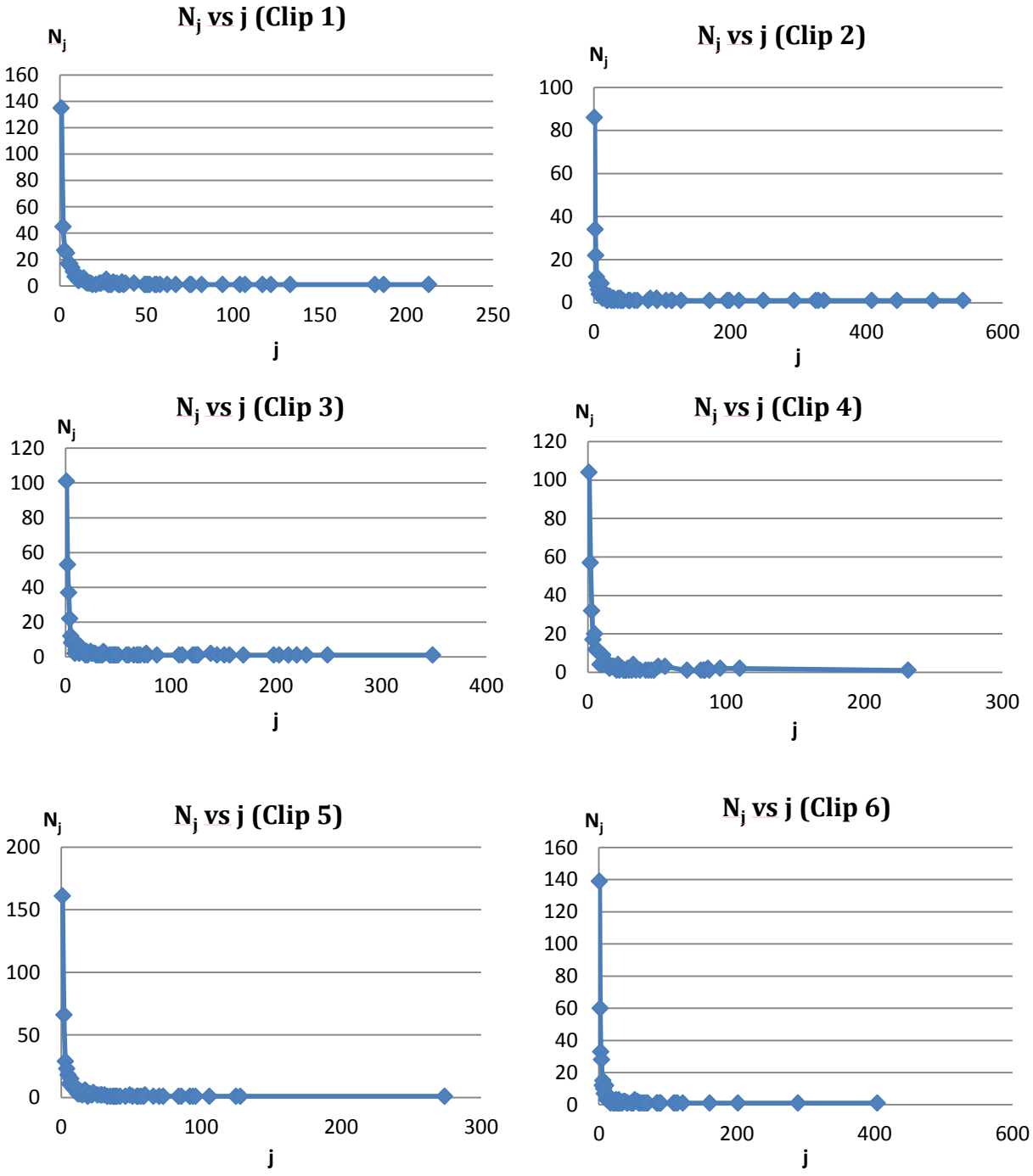


Fig. 5: N_j vs. j plots for six experiment clips

As it is seen, the plots in Figs 4 and 5 follow the respective distribution patterns. Next, we fit the datas into equations (4) to (7), to find T_{MB}, z, α and T_{BE}. The results of the fits are given below in Table 2:

Clip No.	N (total number of note+duration combination)	T_{MB} (K)	z	α	T_{BE}	$\tau = (\ln T_{BE}/\ln N)$
1	4275	1.26 ± 0.05	0.993	1.30	121.49 ± 6.56	0.575
2	7004	0.91 ± 0.04	0.988	0.84	29.90 ± 1.41	0.394
3	6115	1.13 ± 0.04	0.990	1.73	154.22 ± 19.08	0.578
4	3798	1.07 ± 0.02	0.991	1.89	285.87 ± 30.22	0.686
5	4015	1.03 ± 0.03	0.994	1.18	91.60 ± 7.68	0.544
6	4862	0.82 ± 0.03	0.993	0.79	46.07 ± 2.42	0.451

Table 2: Values of the parameters calculated from the experiment

In case of MB distribution, the musical corpus consisting all the note-duration combinations in all the samples has a perfect fit, expectedly, with $R^2 = 1$ (Fig. 2). For the clips separately, the high R^2 values in each of the fitting plots indicate good fitting. While comparing their temperature values, it is seen that clips 1 and 3 has the highest T_{MB} with 1.26 K and 1.13 K respectively. Two of the clips, 2 and 6, recorded T_{MB} values lower than the corpus temperature, i.e., 0.91 K and 0.82 K. Clips 4 and 5 had the closest T_{MB} values to the corpus – 1.07 K and 1.03 K. We compared the temperature observations of the renditions with the observations made by a trained classical singer. This is summarised below in Table 3.

The general trend that can be seen is that the renditions closest to corpus temperature has the most balanced distribution of *Taan*s (the combination of notes rendered in a faster speed) and *Meend*s (a particular technique of Indian music for smooth gliding from one note to other), two of the integral structural units of any vocal performance in Hindustani classical music. Also, proper presence of standing notes is observed in these clips. Whereas, performances having higher T_{MB} have next to no lyrical variations and completely dominated by *Meend*. Contrastingly, lower T_{MB} corresponds to faster note transitions and absence of *Meend*. Hence, T_{MB} is seen to demostrate the kinetic nature of the improvisation of that rendition, quite analoguous to its role in thermodynamics.

Clip No.	T_{MB} (in K)	Observations made by expert
1 3	1.26 1.13	Dominated by *Meend*, no Lyrical variations
2 6	0.91 0.82	Fast note transitions, rhythmic transitions, *Meend* almost absent
4 5	1.07 1.03	Better balance of *Taan*s and *Meend*s. Proper presence of standing notes (*Vadi-samvadi* pairs)

Table 3: Comparing T_{MB} observations with opinions of Classical music expert

Now lets turn to the results of B-E distribution in Table 2. As indicated in [18], the z-value is very close to unity in each of the cases, which is a necessary observation for B-E distribution pattern. In most of the cases, the α values are between 1 and 2. This indicates good fitting for lower j's [20].

The parameter T_{BE} is a measure of the variety in using different notes and their durations, implying higher/lower improvisation pattern with its increase/decrease. Since the B-E distribution implies that low T_{BE} means all the particles in the system have very little energy variations (the lowest T_{BE} being Bose-Einstein condensate where that is ZERO), similarly the lower value of 'temperature' in this case would mean lesser diversity in note occurrences and note durations [21]. Accordingly, clips 2 and 6 display lowest diversity as their T_{BE} are lowest among the six, whereas clip 4 displays highest variety of note usage. The difference of T_{BE} indicates improvisational differences due to singing/performing patterns.

Comparing it with T_{MB} results, we can see that clips that have faster note transitions and absence of *Meend*, also display lesser improvisational variety and clips that are dominated by *Meend* have higher diversity of note usage. This suggests that a clear correlation between presence of *Meend* and improvisation patterns of the performance, as indicated in [22].

The parameter τ (= $\ln T_{BE}/\ln N$) has been observed to be a good variable to look for in the case of comparative studies [19] and the α-τ plane helps in comparisons. In our case, α-τ plane looks like Fig. 6:

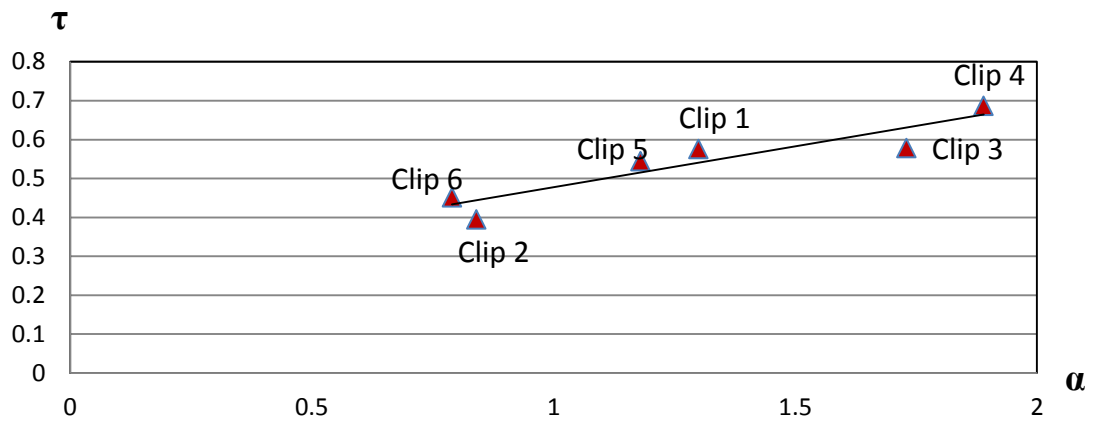


Fig. 6: The position of the clips on the α-τ plane

Higher τ value indicates the existence of lower number of notes used. Evidently, Clip 4 (lying in high α - high τ zone) has fewest and Clip 2 (low α - low τ) has highest number of notes, as seen from Table 2. Also, the spread among notes is seen to be higher in high α-τ zone. Hence, Clips 3 and 4 have less concentrated distribution of notes than 2 and 6, whose distribution of notes are more concentrated (indicates lesser variety of notes used). This observation backs the findings made during T_{BE} results.

6. CONCLUSIONS

The main conclusions that could be summarised from our pilot study are:

1) The fitting of the probability vs energy graphs indicate that even with a small corpus, this method can churn out good results.
2) The parameter T_{MB} demonstrates kinetic nature of the rendition (analogous to thermodynamics) and hence, has the potential to be used as a Classification parameter for musical information research.
3) The consistency in the z-value indicates analogy of B-E distribution and high-structured sequential data congregation such as music is surprisingly significant.
4) Consistency in values of α shows that this method fits well for low frequency data (For high frequencies, spectrum of ε_j needs to be investigated)
5) The parameter T_{BE} can be used to indicate diversity in note variations, and in turn, as an improvisation parameter.
6) There exists a possible correlation between *Meend* and variety of improvisation patterns in the vocal performances of Indian Classical Music; this observation requires further investigations.

7) α-τ plane can be used to categorize different renditions and performance variations for artists based on the nature of note distribution over the whole performance.

Here, we have attempted to present a novel model of investigating the musical information of Indian classical music using fundamental statistical tools (M-B and B-E distributions), extensively used in the domains of physical world. The parameters it yielded can categorize different performances and their improvisational characteristics on the basis of note occurrence and presence of note-duration variation. Usage of such statistical methodologies as a classificatory algorithm in the music domain is unique. With larger data and rendition diversities, further correlation between the parameters and finer categorization of musical information could be possible, we believe. Number of parameters can also be extended to other thermodynamic variables. The early results are indicative that this method could be used in the fields of speech and music for style identification and classification purposes.

7. REFERENCES

1. Muni, M. (1992). Brhaddesi of Sri Matanga Muni. Edited by Premlata Sharma.
2. Gagandeep Hothi; An in depth analysis of Raga Gaavti with new compositions; *Indian Journal of Arts*; 2013; 1(1); 13-16; ISSN : 2320-6659; EISSN : 2320-687X
3. Valla, J. M., Alappatt, J. A., Mathur, A., & Singh, N. C. (2017). Music and Emotion- A Case for North Indian Classical Music. *Frontiers in psychology*, *8*, 2115. https://doi.org/10.3389/fpsyg.2017.02115
4. Gorow, R. (2011). Hearing and writing music: professional training for today's musician. SCB Distributors.
5. Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.
6. Viswanathan, Gandimohan M., et al. "Optimizing the success of random searches." *Nature* 401.6756 (1999): 911.
7. Barabási, Albert-László, et al. "Avalanches in the lung: A statistical mechanical model." *Physical review letters* 76.12 (1996): 2192.
8. Jin, N. Z., Liu, Z. X., & Qiu, W. Y. (2009). Frequency and correlation of nearest neighboring nucleotides in human genome. *Chinese Journal of Chemical Physics*, *22*(1), 27.
9. Zipf, G. K. (1949). Human behaviour and the principle of least-effort. Cambridge MA edn. *Reading: Addison-Wesley*.
10. Mantegna, R. N., & Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature*, *376*(6535), 46.
11. iCancho, R. F., &Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788-791.
12. Gabaix, X. (1999).Zipf's law for cities: an explanation. *The Quarterly journal of economics*, *114*(3), 739-767.

13. Aitchison, L., Corradi, N., & Latham, P. E. (2016).Zipf's law arises naturally when there are underlying, unobserved variables. *PLoS computational biology*, *12*(12), e1005110.
14. Miyazima, Sasuke, and Keizo Yamamoto. "Measuring the temperature of texts." *Fractals* 16.01 (2008): 25-32.
15. Chang, M. C., Yang, A. C. C., Stanley, H. E., & Peng, C. K. (2017). Measuring information-based energy and temperature of literary texts. *Physica A: Statistical Mechanics and its Applications*, *468*, 783-789.
16. Bose, S. N. (1924). Planck's law and light quantum hypothesis. *Z. Phys*, *26*(1), 178.
17. http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/imgqua/disbemb.png
18. Rovenchak, A., & Buk, S. (2011). Defining thermodynamic parameters for texts from word rank-frequency distributions. *Журналфізичнихдосліджень*, (15,№ 1), 1005-1.
19. Rovenchak, A., & Buk, S. (2011). Application of a quantum ensemble model to linguistic analysis. *Physica A: Statistical Mechanics and its Applications*, *390*(7), 1326-1331.
20. Rovenchak, A. (2014). Trends in language evolution found from the frequency structure of texts mapped against the Bose-distribution. *Journal of Quantitative Linguistics*, *21*(3), 281-294.
21. Roy, S., Banerjee, A., Sanyal, S., Ghosh, D., & Sengupta, R. (2019). Categorization of Indian Classical Music Using MB-BE Distributions. *Journal of Software Engineering Tools & Technology Trends*, *6*(3), 9-15.
22. Ghosh, D., Sengupta, R., Sanyal, S., & Banerjee, A. (2018). Improvisation—A New Approach of Characterization. In *Musicality of Human Brain through Fractal Analytics* (pp. 185-212). Springer, Singapore.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

ORNAMENTATION IN HINDUSTANI MUSIC

Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta and Dipak Ghosh
Jadavpur University, India

ARTICLE INFO

Article history:
Received 19/03/2020
Accepted 13/12/2020

Keywords:
Indian Classical Music,
Raga,
Improvisation,
ornamentation,
acoustical analysis

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Most important features in Indian classical music is Melody and it is mainly based on a single raga known as melodic mode. Among a complete raga performance in Indian classical music only a small duration is the pre-composed while the rest involves a gradual systematic exploration of the raga using various forms of improvisation keeping the grammar of the raga intact. The performer has great freedom in exploring his own improvisational patterns in order to delve into the intricacies of the given raga. Artists develop these improvisational patterns based on their experience and expertise. However, the performer is bound by the framework of the raga, tala and other factors such as gharana (the musical school which he belongs to), laya (tempo). There is a great variation in quality and nature of improvisation among different practitioners based on their cognitive and imaginative abilities. Here in this paper we use one of the most important tools used in improvisation is alankar/alankaran (musical ornamentation). All the extempore variations that a musical performer created during a performance within the raga (melodic pattern) and tala limits (rhythmic cycle) could be termed as Alankar or alankaran (ornamentation). These variations in performance embellished and enhanced the beauty of the raga. We are interested in discovering the hidden musical patterns and structures in Indian classical music which are fundamental in unfolding the musical ornamentation of a raga.

1. INTRODUCTION

Indian classical music has traditionally been about single-line melodic development, but artists do very complex things with that one line of melody. Simple melodies are miraculously brought to life through ornamentation (alankar). So, what exactly is ornamentation? The assumption is that there is an underlying skeleton of melody that can stand on its own; ornamentation is what is added to this to make it more appealing. Ornamentation can take place at the level of individual notes – when a note is sung in any manner except straight/steady, it is said to be ornamented. Some ornaments are applied to groups of notes. There can even be ornamentation at the level of the composition itself, such as by introducing melodic or rhythmic variations. Ornamentation or

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Anirban Patranabis
Email:anir.thikana@gmail.com

decorations of a musical performance is an indispensable part of improvisation in Indian classical music which is not allowed in western classical music. The performer tries to evoke the mood or essence of the raga (melodic-mode) with the delicate exploration by flourishing the ornamentation. Unlike western classical music, Indian classical music is cognitively associated with these ornamentations instead of the actual underlying notes only. There are many different kinds of ornaments. Some add finer nuances to the melody, others give it a new dimension in the form of texture. Together, the various ornaments play a very important role in giving body and expressiveness to a simple melody, making it complete in and of itself without the need for accompaniment.

We begin the study with the objective to computationally detect the ornaments in Khayal performances by various eminent vocal singers in Indian classical music. Khayal is one of the oldest surviving form of Hindustani classical music, is monophonic in nature and has a wealth of ornamentation. Here we are interested only the most important form of annotations of primary alankaras (ornamentations), namely, meend (glides between notes) and andolana (oscillations in the note region). This study makes an attempt to study and discuss these ornaments computationally. For this, Sixty-six songs sung by 25 singers covering four ragas namely Bhairav, Todi, Darbari Kannada and Mian-Ki-Malhar were taken for analysis. Only the aalap (only the vocal part of a rendition without any influence of accompanying instruments like table etc.) portion of various ragas sung by various eminent singers in Indian classical music were considered Digitization of the signal was done @44100 samples/second (16 bit/sample).

2. OBJECTIVE

Human brain with little training in Indian classical music can identify style of a singer but an amateur cannot do so. In this work our main objective is to find only the transition between notes (meend) and oscillation (andolan) computationally from the musical signals that help in identifying style of a vocalist.

3. Meend

In Hindustani music, transitional pitch movements between two notes is very important for the development of the aesthetics of a raga. For such a task it is necessary to know what exactly constitutes a cognitive pitch movement as against a note. According to Strawn (Strawn, 1985), a transition "...includes the ending part of the decay (or release) of one note, the beginning and possibly all of the attack of the next note, and whatever connects the two notes." Transitions include a change in pitch, amplitude, and spectrum. In the rendition of Ragas in Indian Classical Music, the used notes, not only their sequences but also the nature of transitions between notes are said to be relevant to characterize the Raga. The role of transitions is also relevant to the concept of ornamentation in Indian music which is essentially to embellish or enhance the inherent beauty of the genre. Such transition between notes are called ‘Meend’. In fact, note transitions play an important role in the domain of Indian Classical Music. Style, emotions, gharana characteristics, raga characteristics and even the personal characteristics might be embedded in these transitions. Therefore, it is very important to measure and classify these transitions (Sengupta R et al, 2006). Some other definitions are as follows.

The Meend in its most basic form can range from a simple span of two notes to a whole octave. These are straightforward, smooth and uni-directional. The basic meend is generally very

slow paces and usually rendered in the first part of the alaap-vistaar. As the pace gradually picks up, the meends also gain in tempo and progress to more complex structures (www.itcsra.org).

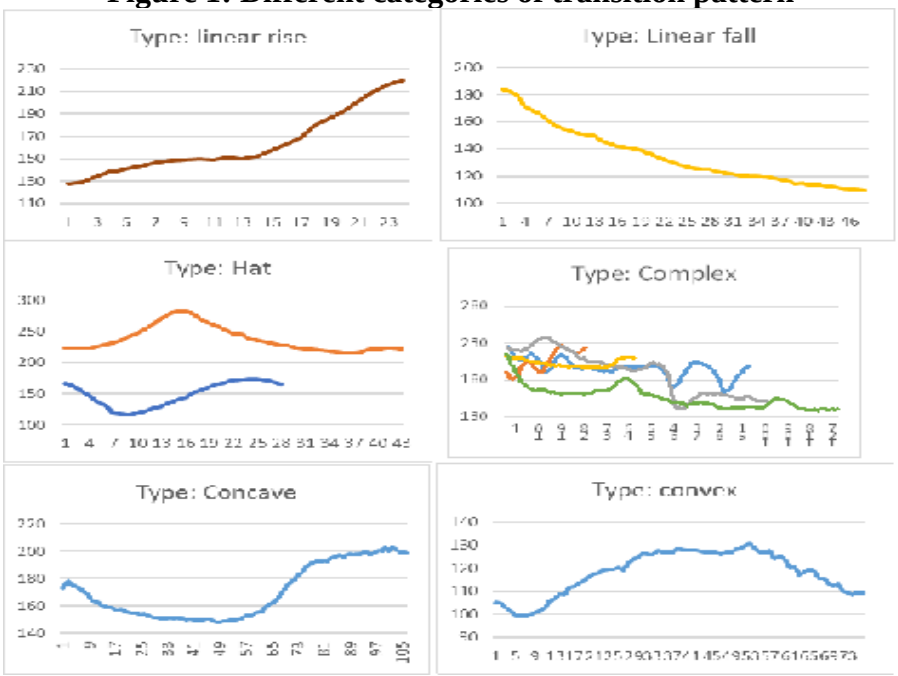
4. Andolan

The andolan alankar is a gentle swing or oscillation that starts from a fixed note and touches the periphery of an adjacent note. What is special about the Andolan is, that in the course of oscillation, it touches the microtones or shrutis that exist in between. The note that is being oscillated within an andolan is known as andolit swar (oscillating note). It must be noted that these andolit swars are raga specific and should not be applied to any raga that uses these notes (www.itcsra.org).

5. Results and discussion

Pitch and hence Meend was calculated as the algorithm prescribed by Datta A K et al 1996, 2008. Figure 1 shows the nature of six different transition categories. In this, the 1st figure shows the pattern of linear rise while the 2nd figure shows the linear fall. Majority of these two types of meend comprise with two to four note transition. The 3rd figure shows the hat category of transition. In this, we found two sub-categories of hat viz. erect hat and inverted hat but we have measured the duration and frequency of occurrences as a whole in a single category 'hat'. In this category, artist glides among three to five notes in which one is at the centre and others are at two adjacent sides of it. In the next figure, we showed only a few among various types of complex pattern of transitions (meend) found in all the renditions. In these kind of meends, various sets of note combinations were used by the artist ranging from four note combination to full octave. Last two figures describe how the concave and convex category of transition looks like. These are smoothly gliding patterns of meend.

Figure 1: Different categories of transition pattern



The numbers of sequences in each of the six broad categories are mentioned in the table 1. We have a data base of three renditions of artist B in the raga Todi, darbari Kannada and Mia-Ki-Malhar while four renditions of artist S in raga Todi, darbari Kannada, Mia-Ki-Malhar and Bhairav. Average duration of transitions (meend duration) in second for six categories by the both the artists sung in those ragas are shown in table 1. Average meend duration is significant in all these categories. For artist B, concave and complex category of meend is highly significant in all three ragas while convex category of meend is least significant. Linear rise and fall durations are also less significant. Long durational concave and complex pattern of transition among notes are important alankaran in the renditions of artist B. Average duration of each category of Hat, linear rise, complex and concave are similarly related among three ragas. This may be a style stamp for this artist. Artist perhaps was not fond of convex type of transition. For artist S, hat or valley category of meend is highly significant in all the four ragas. Linear rise category is also significant in all four ragas. But linear fall category is significant only in raga Todi. Long durational concave, convex and complex pattern of transition among notes are important alankaran in the renditions of artist S. Concave category is significant in Todi while convex category is significant in Mian-ki-malhar. Hat type of meend pattern may be a style stamp for this artist.

Table 1. Average duration of transitions in second for different categories by the artist B and artist S sung in 3 ragas

Meend duration (transition time) of Artist B				Meend duration (transition time) of Artist S			
	Todi	Darbari Kannada	Mia-Ki-Malhar	Todi	Darbari Kannada	Mia-Ki-Malhar	Bhairav
Linear rise	0.272	0.434	0.333	0.709	0.598	0.821	0.748
Linear fall	0.3	0.481	0.132	0.890	0.378	0.546	0.326
Hat	0.564	0.653	0.546	1.112	0.986	1.088	1.376
Complex	0.863	0.667	0.8	0.4	0.785	0.343	0.449
Concave	1.163	1.135	0.898	0.903	0.567	0.490	0.326
Convex		0.09	0.11	0.367	0.639	0.925	0.648

Table 2: distribution of frequency of occurrences of transitions of different categories for different ragas sung by two artists.

Frequency (%) of occurrences for artist B				Frequency (%) of occurrences for artist S			
	Todi	Darbari Kannada	Mia-Ki-Malhar	Todi	Darbari Kannada	Mia-Ki-Malhar	Bhairav
Linear rise	11.11	17.5	23.22	9.8	9.2	6.98	11.43
Linear fall	8.7	7.42	8.66	8.87	9.54	7.51	9.66
Hat	4.32	5.54	1.9	15.5	16.44	12.78	14.62
Complex	10.8	4.76	6.6	4.22	2.55	2.86	2.16
Concave	2.23	2.5	4.4	6.88	9.2	5.55	8.88
Convex		0.5	1.3	6.45	5.89	8.16	6.92

Table 2 shows the distribution of frequency of occurrences of different categories of meend by the two artists. For artist B, frequency of occurrences is significant in all categories of transitions (meend) except for the convex category and for artist S, frequency of occurrences is

significant in all categories of transitions (meend) except for the complex category. Highest frequency of occurrences is observed in the linear rise category of transition for artist B while the same is observed in Hat or valley category of meend for artist S. Linear fall category is significantly similar in all the ragas for both the artists. This may be considered as a style of the two artists. Frequency of occurrences of the meend categories like hat, complex and concave varies differently among ragas for both the artists. This is done by the artist as per the mood or emotion and need to explore the raga. Artist B used a longer span of times in producing concave, complex and linear fall category of meends while most of the linear rise category are of shorter span of time. Artist S used a longer span of times in producing Hat and concave category of meends while most of the linear fall category are of shorter span of time. A negligibly small number of convex pattern was found in all the three signals of artist B. Complex pattern is quite prominent in Todi while linear rise is prominent in darbari Kannada as well as Mian-ki-Malhar for artist B. Undulating fall is seen to be quite prominent in darbari Kannada. Hat pattern is quite prominent in all the four ragas while linear rise and linear fall are prominent in all the four ragas for artist S. This is perhaps a significant style stamp of this artist.

Table 3. Average duration of transitions in second and frequency (%) of occurrences of transitions of four ragas for 19 artists

Artist	Bhairav	Darbari	Mia-ki-Malhar	Todi
A		0.507 , 4.97		
Bak			0.335 , 4.04	
Bal			0.578 , 6.04	
Bg		0.486 , 5.25		
Bh		0.577 , 6.37	0.47 , 7.68	0.632 , 7.43
D		0.597 , 6.05		
F				0.439 , 4.67
Ga			0.458 , 5.21	
Gi	0.696 , 7.08	0.625 , 6.95		
J			0.562 , 4.86	0.456 , 5.17
Ka			0.761 , 6.98	
Ki	0.746 , 6.88			
Lat				0.821 , 8.67
Lax	0.803 , 7.88		0.651 , 7.34	
Mr			0.452 , 5.13	
Mu		0.556 , 6.05		
N			0.56 , 6.87	
Sal	0.646 , 8.95	0.659 , 8.8	0.702 , 7.31	0.73 , 8.62
Sar		0.445 , 5.63		

Now we have a database of twenty-seven renditions of four ragas namely Bhairav, darbari, Mia ki Malhar and Todi sung by nineteen artists. Average duration of transitions in second and frequency (%) of occurrences of transitions of 27 renditions are shown in table 3. It is observed that the variations in average meend duration and frequency of occurrences of meend among artists are quite less in raga Bhairav and Darbari while this variation is significantly higher among artists for rags mia ki Malhar and Todi. In general, artists prefer quite long durational meend in raga Bhairav but no such preferences are observed for other ragas. Minimum average

meend duration observed is of 0.335 second. So Meend in Hindustani Music is undoubtedly an indispensable ornament.

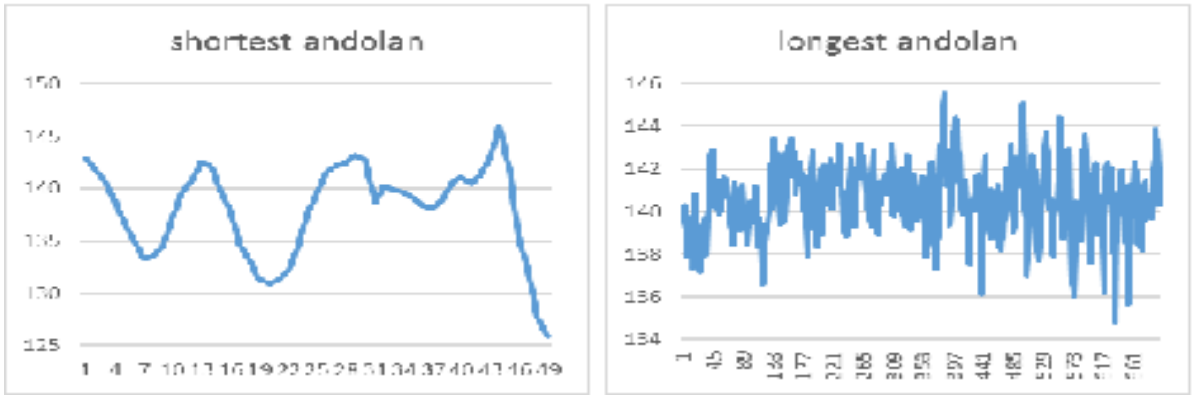


Figure 2: Shortest and longest andolan type (both are about komal re and from raga bhairav)

Table 4: Andolit Swaras (notes about which the oscillation takes place)

Bhairav	Darbari	Mian-ki-Malhar	Todi
Komal (flat) re	Komal (flat) ga	Ni	Komal (flat) ga
Dha	Komal (flat) dha	Komal (flat) ga	Dha

Table 5: Average andolan duration in milliseconds and frequency (%) of occurrences of andolan

Artist	Bhairav	Darbari	Mia-ki-Malhar	Todi
A		182.3 , 3.28		
Bak			292.3 , 1.49	
Bal			144.1 , 1.94	
Bg		297.4 , 3.81		
Bh		292.8 , 0.87	301.5 , 2.53	381.9 , 3.37
D		317.4 , 4.39		
F				323.1 , 4.13
Ga			310.6 , 1.84	
Gi	195.5 , 7.44	243.9 , 1.89		
J			296.3 , 3.03	280.6 , 8.97
Ka			352.1 , 2.48	
Ki	464 , 6.14			
Lat				187.8 , 11.11
Lax	236.3 , 4.1		162.2 , 2.25	
Mr			282.9 , 3.77	
Mu		212.6 , 2.76		
N			324 , 3.2	
Sal	224.7 , 5.63	304.3 , 4.68	259 , 3.55	315 , 4.17
Sar		457 , 3.86		

Figure 3: Comparison in andolan for two artists

6. Conclution

In the context of Indian classical music, the application of note transition and oscillation about a note plays an important role in embellishing inherent beauty of the genre. These two are very important style stamp of a vocalist. This style characteristic is well understood by different intonation pattern like flat, rise, fall, hat, valley etc. In Hindustani music artists have enough liberty to improvise them performance within constraint of fixed framework of a given raga. Ornamentation or alankar is a part and parcel of improvisation in HM. Proper and useful utilisation of alankar bring appreciation from listeners. Simple flat meend with 2 or 3 note combinations or complex type of meend with more notes combination evokes different emotion to the listeners. The type of meend is an unique identification of style of a performer. Some prefer short but repeated use of meend while some prefers long durational slow tempo meend. Duration and frequency of andolan is also a style stamp of an artist. This study will be further extended to find other alankaran cues like gamak, murki and khatka. This will lead to distinguish the style of vocalists.

7. References

Strawn J, "Modeling musical transitions," Ph.D. dissertation, Stanford University, Stanford, CA, 1985.

Sengupta R., Dey N. and Nag D., Extraction and Relevance of Transitory Pitch Movements in Hindusthani Music, Proceedings of the NSA-2006, N. Delhi

Datta A K, Sengupta R, dey N and Nag D, A methodology for Automatic Extraction of ‘Meend’ from the Performances in Hindustani Vocal music, Journal of ITC Sangeet Research Academy.

Datta A K, Generation of Musical Notations from Song using State-Phase for Pitch Detection Algorithm, Journal Acoustical Society of India, Vol XXIV, 1996.

Datta A K, Sengupta R, dey N and Nag D, Automatic Classification of ‘Meend’ Extracted from the Performances in Hindustani Vocal Music, Proceedings of FRSM -2008.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

A InTraSAL (Intonational Transcription of South Asian Languages) analysis of Standard Colloquial Bengali

Moumita Pakrashi, Shakuntala Mahanta
IIT Guwahati, Guwahati, Assam, *India*

ARTICLE INFO

Article history:
Received 13/12/2020
Accepted 22/12/2020

Keywords:
spontaneous speech,
intonation,
InTraSAL,
pitch accents,
boundary tones,
focus

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU

Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

In this paper a pilot study of spontaneous speech Bengali data has been carried out using the recently developed phonological model and annotation system of **InTraSAL**. Majority of the intonational studies on Bengali speech are based on scripted and readout sentences. However this study aims at eliciting spontaneous speech in order to study its intonational pattern and observe how much that varies from the scripted speech analysis patterns. Based on InTraSAL analysis, it can be seen that Bengali bitonal pitch target points High(**H**) and Low(**L**) has varied patterns of pitch accents and boundary tones in different sentence types and focus structures.

1. Introduction

Ladd defines intonation as '*…the use of suprasegmental phonetic features to convey postlexical or sentence-level pragmatic meanings in a linguistically structured way.'* (Ladd 2008:p4) Intonation is studied in several areas of linguistics, including pragmatics, semantics, syntax, phonology and phonetics. It has been a major area of research in speech for quite some time now. The significance of a precise intonational analysis is furthermore widely acknowledged in areas of speech technology, particularly in speech synthesis and speech recognition systems.

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Moumita Pakrashi
Email: mpakrashi1@gmail.com

Bengali is a part of the Indo-Aryan branch of the Indo-European group of languages. It is the official language of the Indian state of West Bengal and the national language of Bengalidesh. It is also spoken in the adjacent states of Assam, Tripura and so on. Bengali is the fifth most-spoken native language in the world

This paper studies the intonational phonology of **Standard Colloquial Bengali** as spoken mostly south-east part of West Bengal and specifically in and around Kolkata. Data used in this paper is spontaneous Bengali speech. It has been elicited through narrative descriptions by the subjects. The phonological analysis has been done using the **Intonational Transcription of South Asian Languages (InTraSAL)** model. This is a phonological model and annotation system that is being currently used to analyse the intonation of a range of South Asian languages with similar prosodic characteristics. Sameer ud Dowla Khan had first proposed this model for Bangladeshi Bengali (as B-ToBI) in his dissertation. InTraSAL has been developed on the framework of autosegmental-metrical (AM) theory of intonational phonology and the ToBI-style method of prosodic annotation. AM theory of intonation and the ToBI was primarily developed by Pierrehumbert (1980) and Beckman. Preliminary works of intonational study of Kolkata centric Bengali was done by Chatterji (1921), Ferguson & Chowdhury (1960), and Ray, Hai, & Ray (1966). Later on based on AM model of intonation, Bengali prosody was studied by Hayes & Lahiri (1991), Lahiri & Fitzpatrick-Cole (1999), Michaels & Nelson (2004), Jun (2005), and Selkirk (2006). The first ToBI transcription system of Bengali was proposed in Michaels & Nelson's (2004) model.

2. Data Collection

Data was collected from 7 speakers of Kolkata standard Bengali, who are natives of Kolkata and surrounding areas. Of them 3 were male and 4 female. Time span of the entire audio data was approximately of 51minutes. Data collection for spontaneous speech was done by asking speakers to narrate stories in Bengali by providing them two types of stimulus. First, a silent video clip known as "The Pear Story" was shown to them and asked to narrate the event in their own words in Bengali. Next some pictorial representations of stories were presented, and speakers were asked to reproduce the story. The elicitation using pictures was done with the help of 'Totem Field Storyboards'. The storyboards combine the advantages of eliciting spontaneous speech with expressive content along with different syntactic elements, focus and sentence constructions. The recording was done using a Tascam DR-100 MKII and the analysis was done using PRAAT software (Boersma and Weenink 2015). The annotations were all done manually.

3. InTraSAL analysis of Standard Colloquial Bengali

In this paper, various tonal patterns in Bengali speech are analysed in Praat software. Praat pictures used below to illustrate each tonal pattern show the image of pitch contour of the speech segments in blue. Below the pitch contour are three transcription tiers:

• The tones tier shows the tonal events, e.g. L*, Ha

• The IPA words tier shows segmentation and IPA transcription of the text, e.g. 'koɪt͡ʃʰe'
• The English tier shows their corresponding English words and suffixes, e.g. ' DO-PRG-3'

3.1 Prosodic Structure

The analysis done using InTraSAL on Standard Colloquial Bengali describes three types of prosodic units. They are Accentual Phrases (AP), Intermediate Phrases (ip), and Intonational Phrases (IP), which are all composed of two basic pitch accents – Low(L*) and High(H*) and several other boundary tones. The smallest prosodic unit above the word level is an AP followed by an ip and finally the largest being an IP.

a. Accentual Phrase

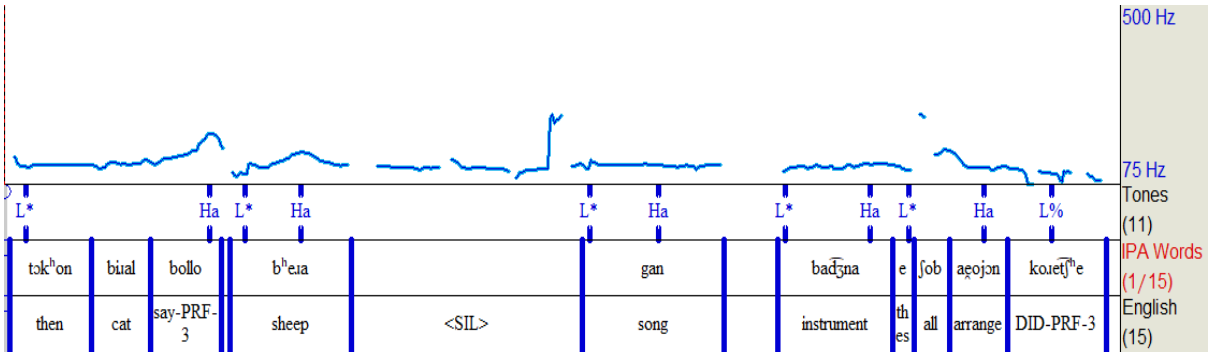
Here the basic unit of prosody is the accentual phrase (AP), which is underlyingly composed of exactly two tones: a pitch accent (low L* or high H*) and an AP boundary tone (high Ha or low La). Pitch accents are tones that highlight the most prominent syllable of a word; in Bengali this is usually the word-initial syllable. The right edge of the AP is marked by the boundary tone (Ha/La).

Prenuclear Accentual Phrase

One pitch accent and one AP boundary tone make up a prenuclear AP, which is almost every word and comes in two types: rising (L*...Ha) and falling (H*...La).

Rising AP (L*…Ha)

In prenuclear APs, pitch accents can be either high (H*) or low (L). The rising AP (L*..Ha) is most common pattern in Bengali.

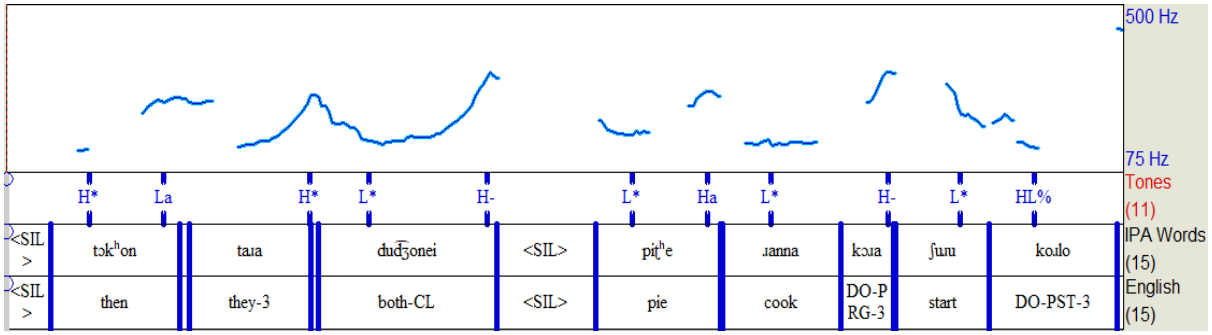


"Then cat said, 'Sheep has organised songs and music.'"

Fig.1: The APs have pitch accents (L*) with their opposite boundary tones (Ha).

Falling AP (H*…La)

Only sporadic uses may be noted of this AP pattern.



"Then they both started baking pies.'"

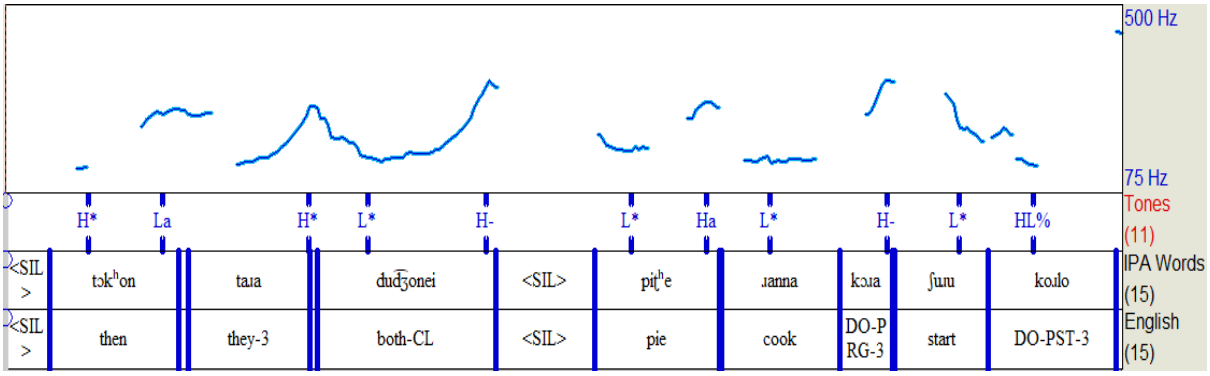
Fig.2: The first AP has pitch accent (H*) with their opposite boundary tones (La).

Nuclear Accentual Phrase

After a sequence of prenuclear APs comes the nuclear AP, which is composed of only a pitch accent, and no AP boundary tone. Three nuclear pitch accents are low (L*), high (H*), and rising (L*+H). This is the final AP within the sequence.

Low pitch accent (L*) and high pitch accent (H*)

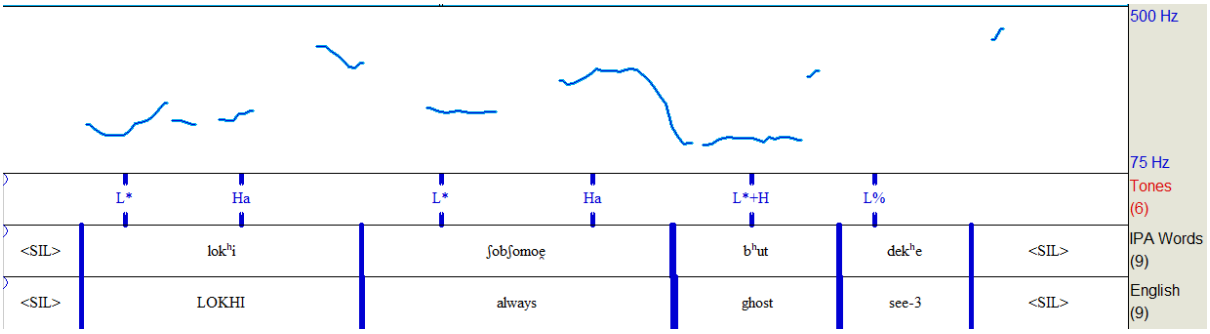
The low pitch accent is default in a nuclear AP. But high pitch accent is also quite common found in the main verb of the sentence. The example below shows both the low and the high pitch accent found in nuclear APs in the same sentence.



"Then they both started baking pies.'"

Fig.3: The two nuclear APs have high pitch accent (H*) on the subject here 'they', and low pitch accent (L*) on the main verb 'start'.

Rising pitch accent (L*+H)

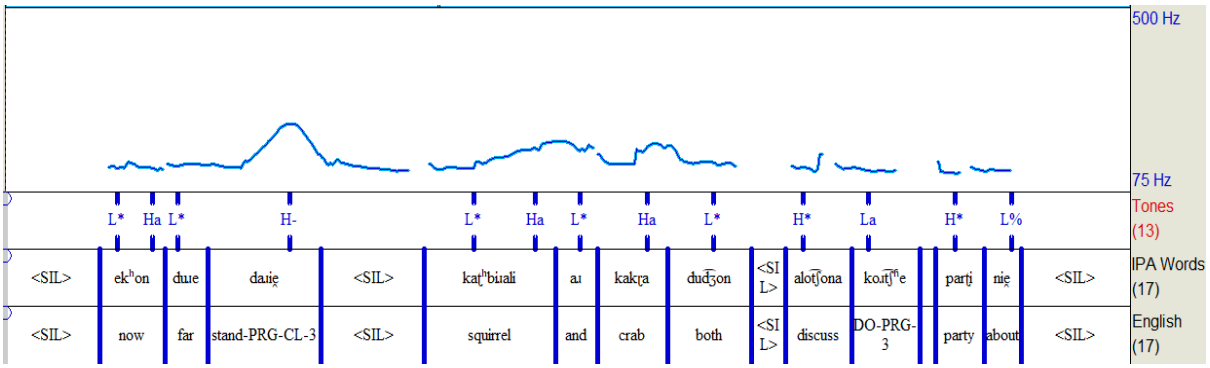


"Lokhi always sees ghosts.'"

Fig.4: The nuclear AP here has a rising pitch accent (L*+H) on the object here 'ghost'.

b. Intermediate Phrase

The intermediate phrase is a cluster of one or more APs, mostly representing a tight syntactic unit, e.g. a topic, subject, or postpositional phrase. Its right edge marks lengthening of the final syllable, an optional pause and one of two boundary tones: high (H-) and low (L-). High ip boundary tone has been commonly found in our analysis.



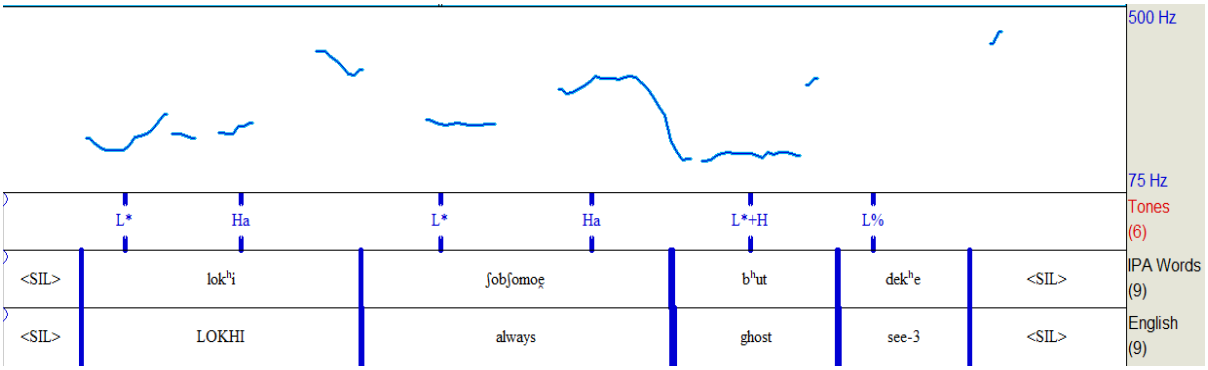
"Now, standing afar squirrel and crab were discussing about the party."

Fig.5: The high (H-) ip boundary pattern is commonly found in declarative sentences.

c. Intonational Phrase

The intonation phrase (IP) is the largest unit that is marked by intonation. Whole sentences can be a single IP, but smaller chunks (even a single word) can serve as an IP as well. One or more ips are grouped together to form an IP. The right edge of an IP is marked by one of the four boundary tones—low (L%), high (H%), falling (HL%), and dipping (HLH%)—which override the boundary tones of the IP-final ip. The IP boundary tone may be H% or L% primarily depending upon the sentence type i.e declarative or yes-no question.

Low IP boundary tone (L%)

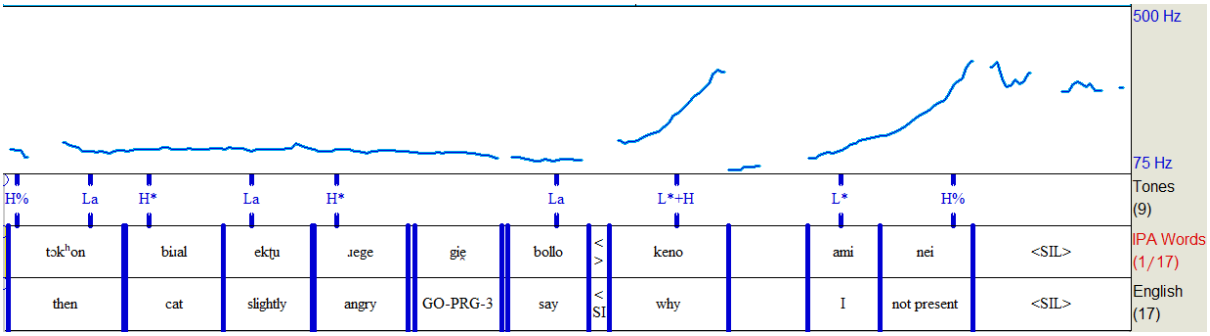


"Lokhi always sees ghosts.'"

Fig.6: The most common IP boundary tone here is the low tone (L%), found in declarative sentences.

High IP boundary tone (H%)

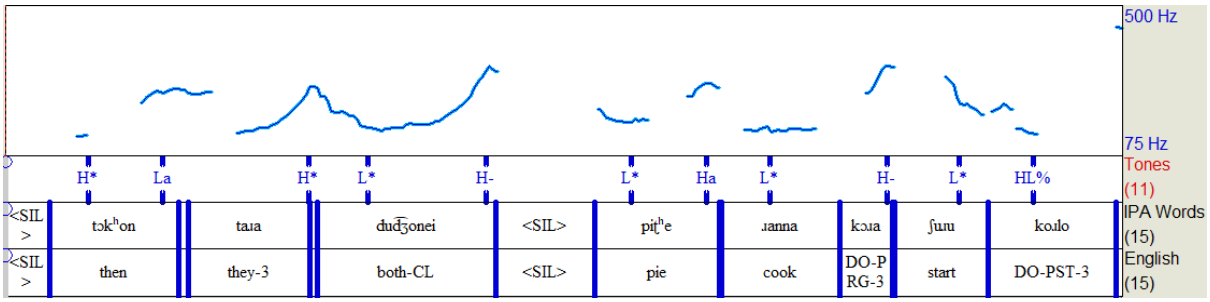
Fig.8 illustrates the high IP boundary tone (H%) usually found at the end of interrogatives.



"Then getting irritated the cat said,' Am I not here?"

Fig.7: The IP boundary tone for yes-no question are H%.

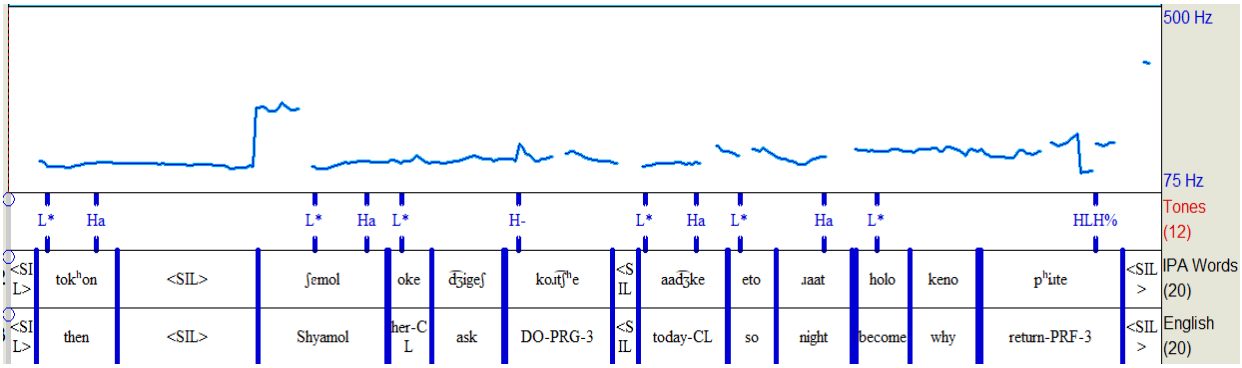
High falling IP boundary tone (HL%)



"Then they both started baking pies.'"

Fig.8: The final high falling IP boundary tone (HL%) marks the final topicalized phrase ʃuɹu koɹlo 'to start'.

High dipping IP boundary tone (HLH%)



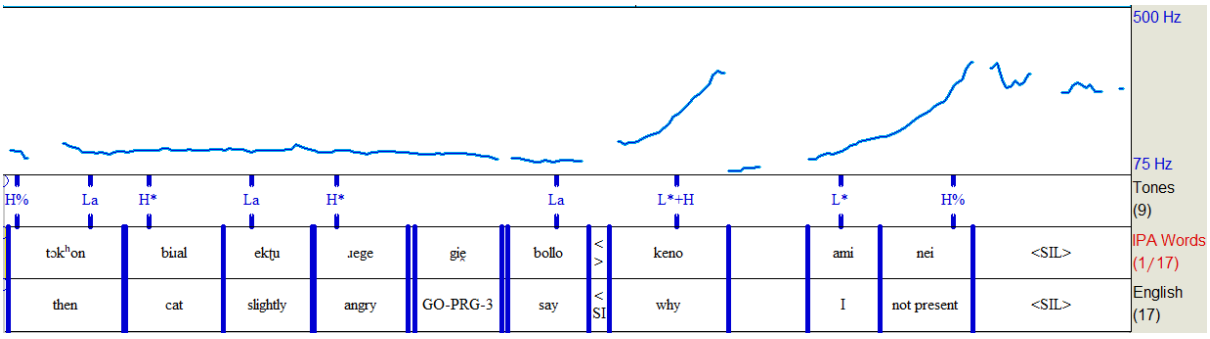
"Then Shyamol asked her, 'Why were you so late tonight?'"

Fig.9: The Wh-question shows a tritonal IP boundary pattern, the dipping (HLH%).

The dipping IP boundary tone (HLH%), made up of three targets is composed of two H targets separated by an L target. Khan (2014) mentions that (HLH%) is used on non-sentence-final phrases, and especially non-final dependent clauses: relative clauses, because-clauses, if-clauses, etc. But surprisingly here it is found in the IP final position of a Wh-question. It is realized as rising pitch beginning from the pitch accent and ending at the boundary between the penultimate and final syllables, followed by both a fall and a rise in pitch during the IP final syllable.

d. Obligatory Contour Principle(OCP)

As shown by Hayes and Lahiri, Bengali intonational contours are indeed governed by the Obligatory Contour Principle (OCP), that prevents the occurrence of adjacent identical tones in intonation contours. Though it is a majority rule, there are a few exceptions to that as:



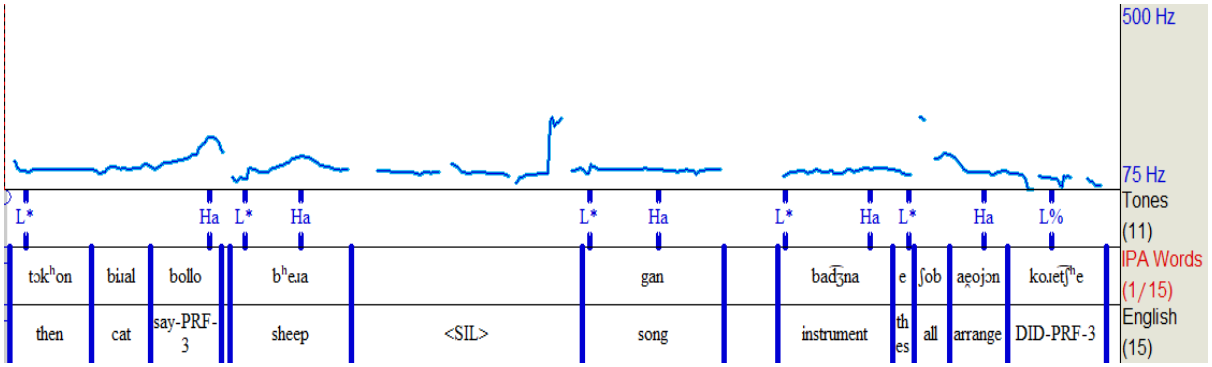
"Then getting irritated the cat said,' Am I not here?"

Fig10: The falling APs (H*..La) gradually change to rising pattern (L*..Ha)'.

3.2 Sentence types

Declaratives, imperatives and interrogatives end with IP boundary tones that are dependent on the particular sentence type. All the five IP boundary tones (i.e. L%, H%, HL%, LH%, HLH%), are used at the end of a complete sentence.

a. Declaratives
Mostly declaratives are marked by the low IP boundary tone (L%) as in Fig.12.

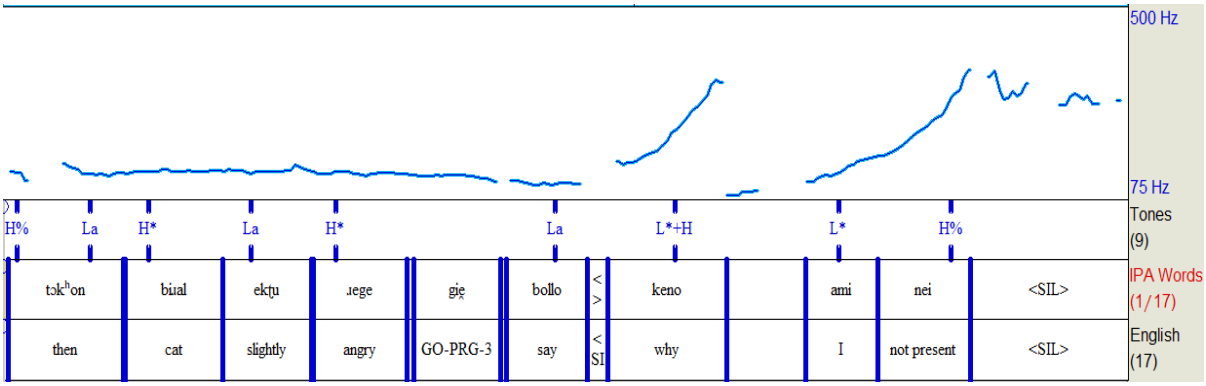


"Then cat said, 'Sheep has organised songs and music.'"

Fig.11: Most declaratives have a low IP boundary tone(L%).

b. Interrogatives

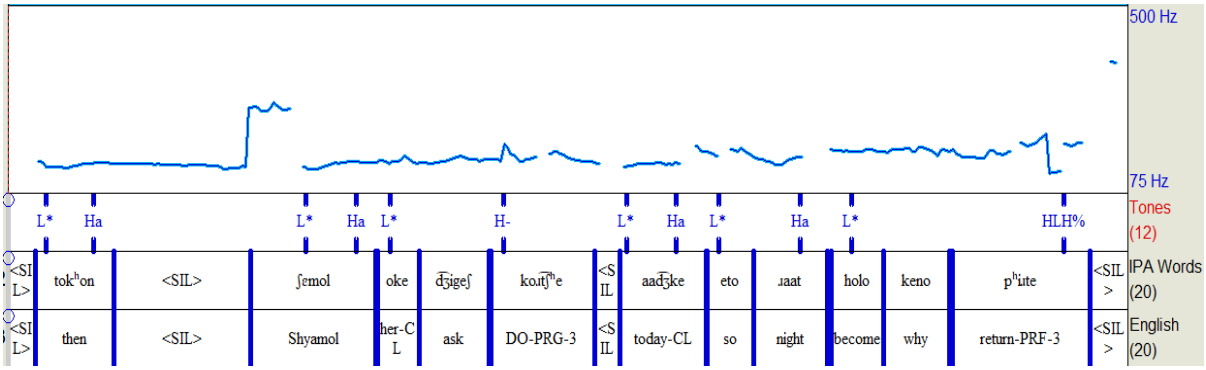
Yes-No questions:



"Then getting irritated the cat said,' Am I not here?"

Fig.12: The usual high IP boundary tone (H%) usually found at the end of yes-no interrogatives..

Wh-questions:

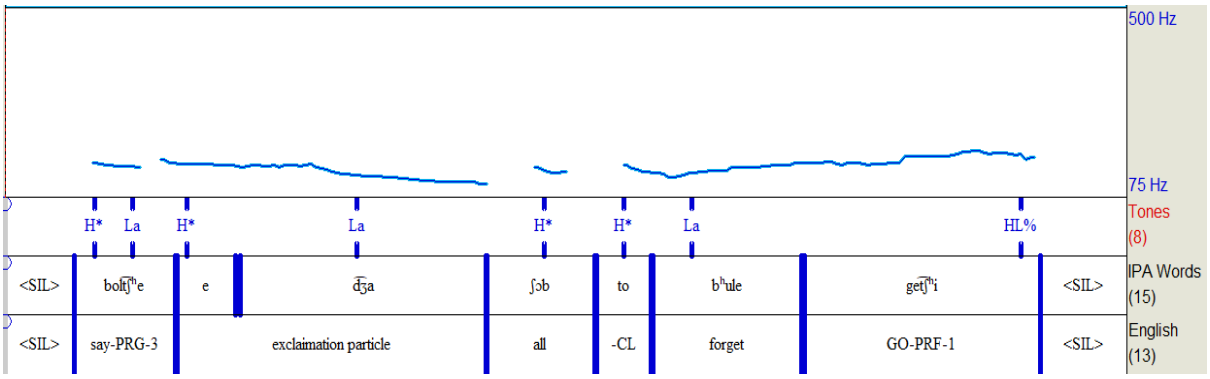


"Then Shyamol asked her, 'Why were you so late tonight?'"

Fig.13: The Wh-question shows a tritonal IP boundary pattern, the dipping (HLH%).

Yes-no and Wh-questions, the two broad types of interrogative sentences bear two types of IP boundary tones : H% and HLH% respectively.

c. Exclamatory



'Says,"Oh no! I forgot everything.'"

Fig.14: The final high falling IP boundary tone(HL%) marks the exclamation tone of the sentence'.

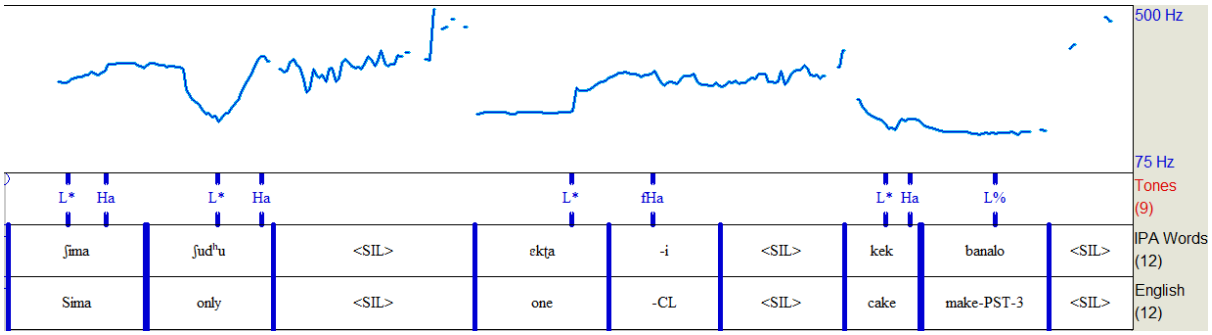
Thus exclamatory sentences have a rising IP boundary tone (H%), just like interrogative sentences.

3.3 Focus realization

Focus is realized prosodically in the use of a special high tone (fH). The realization of this abstract tone depends on the type of focus (i.e. corrective, encliticized, or surprise). The focused high tone (fH) helps the focused element stand out as the most crucial part of the sentence. Three different fH realization patterns are seen in three different contexts. The fH tone attaches to an AP-level tone, i.e. a pitch accent or AP boundary tone.

a. Encliticized Focus

Enclitic markers are added as affixes at the end of independent words. In Bengali the focused rising AP (L* . . . fHa) is used on words with focus enclitics -[i] "only" or -[o] "also," "even," which attach directly to the right edge of the word under focus.

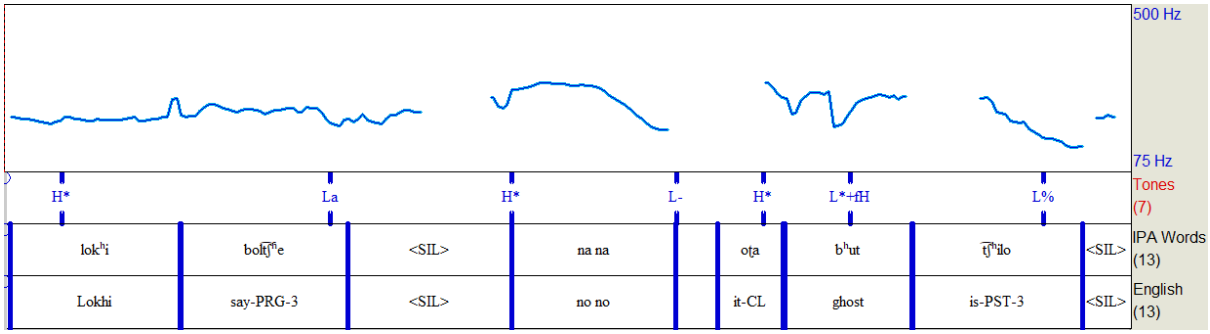


"Shima made only one cake."

Fig.15: Sentences bearing focus markers through enclitics like –[i] or –[o], receive the focus high tone (fH) on the PP boundary tone (Hp).

b. Corrective Focus

In corrective focus, focused constituents are found as corrections to inaccurate statements as the name suggests. They bear the focused rising pitch accent, composed of a single pitch accent with two tonal targets (L*+fH). This pitch accent appears as an F0 valley on the focused constituent.

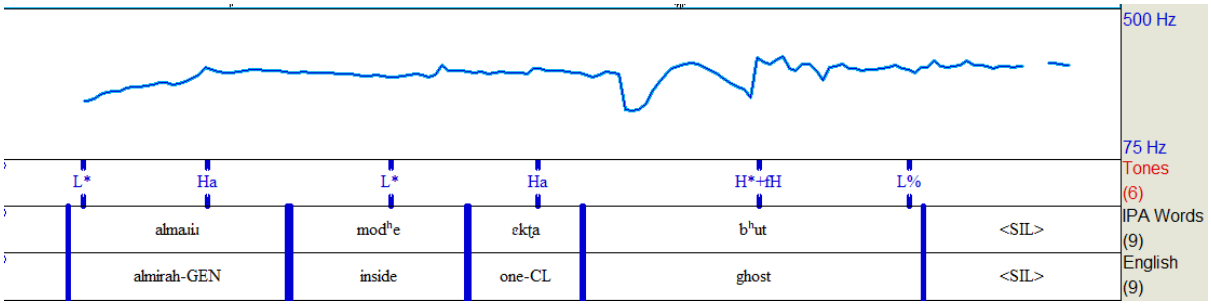


"Lokhi said that 'no, no, it was a ghost'".

Fig.16: Sentences with corrected focused constituents indicating [bʰut] in this example bear the focused rising pitch accent (L*+fH).

c. Surprise Focus

The unexpected information in such a sentence often creates a sharp rising AP as H*+fH preceeded by a high pitch. In surprise focus the focus high tone (fH) fuses with the previous high pitch accent (H*) to form a focused high rising pitch accent (H*+fH).



"In the almirah there is a ghost!"

Fig.17: Statements with surprise focus, here on the word [bʰut]) bears a focused high rising pitch accent (H*+fH).

4. Conclusion

The InTraSAL analysis of Standard Colloquial Bengali finds three prosodic units of Accentual phrase (AP), Intermediate phrase (ip) and Intonational phrase (IP). There are two pre-nuclear AP tones (L*..Ha and H*..La), three nuclear AP tones (L*, H* and L*+H), two ip boundary tones (H- and L-), four IP boundary tones (L%, H%, HL%, HLH%). Furthermore, all AP tones (i.e. pitch accents and AP boundary tones) are more or less guided by the OCP constraint to bear opposite tonal targets; though there are exceptions to the rule.

Also the underlying focus high tone (fH), are expressed in three different manners depending on the type of focus applied. The focus high tone (fH) fuses with the high AP boundary tone (Ha) in encliticized focus constituents, fuses with the high pitch accent (H*) in surprise focus constituents, and joins itself with the low pitch accent (L*) in corrective focus constituents. Unlike previous studies of Bengali intonation patterns, the aim of this study was to focus on the spontaneous speech data rather than read out speech. But in this study of association between intonational tones and spontaneous speech for Standard Colloquial Bengali, it can be found that most of the findings comply with that of Hayes & Lahiri work on Bengali scripted speech as well as Khan's work on the Bangladeshi variety.

Being only a pilot study of Bengali spontaneous speech, the findings of this paper would require larger amount of data study in order to establish them. Further the InTraSAL model is in the process of further development, which means there might many more additions of parameters to the model. The model is also being applied to many more South Asian languages in order to find more similarity in the intonational patterns of these geographically bound languages. Such studies of multiple languages in future would always enable to renew the dimensions of intonation study in Bengali.

References

Chatterji, Suniti Kumar (1921). 'Bengali Phonetics'. Bulletin of the School of Oriental Studies. University of London.

Hayes, B., & Lahiri, A. (1991). Bengali Intonational Phonology. Natural Language and Linguistic Theory.

Khan, S. D. (2010). Bengali (Bangladeshi Standard). Journal of the International Phonetic Association.

Khan, S. D. (2014). The intonational phonoloy of Bangladeshi Standard Bengali. In S.-A. Jun, Prosodic Typology II.

Khan, S. D. (2008). 'Intonational Phonology and Focus Prosody of Bengali', Ph.D. dissertation, University of California, Los Angeles.

Ladd, D. Robert (2008). Intonational Phonology. New York: Cambridge University Press.

Ray, Punya Sloka, Muhammad Abdul Hai, & Lila Ray (1966). Bengali Language Handbook. Center for Applied Linguistics. Washington, D.C

Jun, Sun-Ah (ed.) (2005). Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press.

www.pearstories.org

www.story-builder.ca

www.reed.edu/linguistics/khan/B-toBI/

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Cognitive Functions in Non-musicians in Music Perception

Sukdeb Goswami
Utkal University of Culture

ARTICLE INFO

Article history:
Received 08/04/2020
Accepted 05/01/2021

Keywords:
musical syntax
multisensory perception
cognition of spatial abilities
task switching

Guest Editors:
Dipak Ghosh
Shankha Sanyal
Pijush Kanti Gayen
Ratul Ghosh

Organized by
School of Languages and Linguistics, JU and Centre for Physics and Music, JU
Supported by
JU RUSA 2.0
SERB, DST

ABSTRACT

Music is generally considered to be a complex, auditory, non-verbal material that has syntax of its own. Such syntax is the structural organization of musical events or sub-sets developed over time. Listening to music demands certain perceptual abilities like pitch discrimination, auditory memory, attention, temporal and harmonic patterns of musical syntax etc. Hence the auditory cognitive system must depend on working memory mechanisms. Such mechanisms allow perceptual simulations to maintain relationality with one element in a sequence to other elements in subsequent sequences. My study aims at observing how this cognitive process in non-musicians increases communication co-ordination, cooperation and empathetic attitudes among in-group members. It will also highlight how multisensory perceptual feedbacks are necessary to understand tonal regularities and deviations and how listening to music becomes context dependent experiences like verbal speech acts. In my study I shall mainly deal with the ideas related to acquisition of spatial abilities in language acquisition and domain specific task switching that is a component part of musical syntax processing. Here the crucial methodological assumption is that both verbal and non-verbal language reflects thought process. The investigation process will mainly be based on general scientific methods like observation, induction, deduction, analysis, synthesis, interpretation as well as principles of anthropocentrism that will lead to a structural-functional integrity.

1. Introduction:

Listening to music is a pleasurable journey. Any new experience to musical exposition requires some initial adjustment and insight that may be either very exciting or unbearable depending on

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Sukdeb Goswami
Email: sukdeb99@gmail.com

a varied number of perspectives. But with a little preparedness and insight, listening to music may become a potentially rewarding experience. It is claimed that humans have innate predisposition for music. The claim is mainly based on researches on music conception abilities in infants. Even a non-musician individual has ability to detect when someone sings out of tune, to recognize a familiar tune and to recognize even short melodies. From ancient times listening to music is an institutionalized cultural practice and subject of theorization. In sixth century B.C. in Greece Pythagoras founded a model based on simple string-length ratios that formed the consonances of the Octave. This is noted as "Pythagorean tuning" that reigned in musical theory up to the beginning of sixteenth century. Other theorists like Ptolemy and Boethius contributed to this division of *'tetrachord'* by two intervals (namely *'Semitonium'* and *'Tonus'*). Up to the 19th century musical theories were mainly based on mathematical modeling of pitch and tone. It is believed that Aristoxenus foreshadowed modern music psychology. But from the vast literature of musicology one thing is evident that simple ratios or mathematical modellings are not enough to account for musical phenomenon and a perception based cognitive approach is necessary. In this paper I have aimed at studying the cognitive functioning in non-musicians or persons who have very little knowledge in elemental properties of music. This paper will mainly deal with the research question: *"How do the multi-sensory perceptual feedbacks and context dependent experiences coordinate in music perception by the non-musicians?* To justify the rationale of my study, I have mainly dealt with the theories of perception based cognition. I have also considered the percepts of songs as well as instrumental music perception to reach to the conclusion.

2. Basics of Musical Syntax:

Listening to music is a multisensory-motor experience in which complex symbolic signals are transformed and synthesized into sequential, bimanual, motor activity that depends on multisensory feedback. It also requires precise timing of several hierarchically organized actions and control over pitch interval production (Zatore, 2003). Basic elements of music include ***'temporal elements'*** like rhythm, meter, tempo, dynamics, accent etc and ***'pitch elements'*** like timbre, melody, harmony etc. Any music theory basically focuses on melody and harmony. Temporality is a very important factor in music perception since the placement of the sounds in a temporal sequence is the *'rhythm'* of a piece of music. Hence *'rhythm'* is simply a placement in time. *'Beat'* refers to a specific repetitive rhythmic pattern that maintains the pulse. Beats are naturally a grouping into measures or bars. The first beat tends to be the strongest and generally most of the bars have the same number of beats. The underlying patterning in the pulse of music is established by these things. On the other hand, *'meter'* in music is the arrangement of rhythms in a repetitive pattern of strong and weak beats. Meter is a very useful way to organize music. Another important term in music perception is *'tempo'* which reflects how fast the music should feel. It depends on several things like texture, complexity of the music, how often the beat gets divided into faster notes and how fast the beats themselves are. Gradual change in basic tempo is a very common event in music. *'Dynamics'* is a relative feature of sound that may be barely audible or very much loud even to hurt our ears. *'Accents'* in music are markings that are used to refer to strong sounding notes. Both *'dynamics'* and *'accent'* depend on the instrument playing it as well as on the style and period of the music.

Among pitch elements in music, timbre deals with those aspects of a musical sound that does not have anything to do with the sound pitch or loudness or length. If a same note is played in 'Sarod' and 'Santoor' for the same length of time and in same loudness, the difference between sounds of two instruments can easily be distinguished in relation with the timbre of the sounds. Timbre is created because each note from a musical instrument is a complex wave containing more than one frequency. The human ear and brain can hear and distinguish very small vibrations in timbre. Thus same musical instrument played by two different artists can be distinguished with the knowledge of timbre. Another basic pitch element in music is *'melody'*. *Melody* is not merely string of notes. The line that sounds most significant in a note is the melody. When melody progresses, the pitches may rise or fall slowly and quickly. There is a great similarity between a grammatical phrase in a sentence and a melodic phrase in music. A grammatical phrase is a group of words that make sense together and can express a specific notion; still it is not complete as in the sentence: *I am under* ***the heavy weight of a dream.*** Similarly, a melodic phrase is a group of notes that makes sense together and expresses a definite melodic 'notion', but it is not complete until it accommodates a number of phrases together. *'Harmony'* in music is the result of more than one pitch sounding at the same time in music. Thus, when there is more than one pitch sounding at a given time, there is harmony. An organization of all these elements in a systematic structure forms the semantic property of musical syntax. Thus the semantics of music begins with a large number of semiotic properties described so far.

3. Neurological or anatomical bases of music perception:

When the auditory stimulus enters the ears, it undergoes the processes of the ears. Then it enters the auditory cortex which is part of the temporal lobe. It begins processing the sound by assessing its pitch or volume. This time the brain functioning differs among the analysis of different aspects of music. For example, the rhythm is processed and regulated by the left frontal cortex, the left-parietal cortex and the right cerebellum standardly. Tonality is assessed by the prefrontal cortex and cerebellum (Abram 2013).

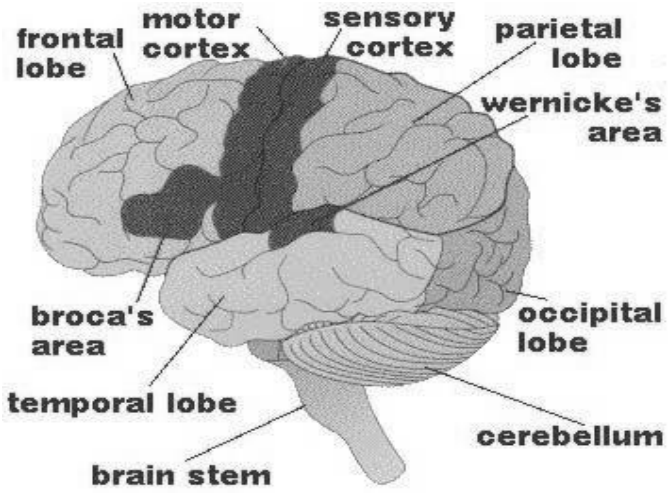


Figure- 1; Simple Brain Diagram

The *Broca's* and *Warnicke's* areas known for activating during speech and language processing, are found to be activated while listening to unexpected musical chords. It establishes the relationality between language acquisition and music perception. In a study Richard Kunert, Roel M. Willems, Daniel casasanto, Anirudh D. Patel and Peter Hagoort used functional magnetic resonance imaging (fMRI) in conjunction with an interference paradigm based on sung sentences. In this study the researchers showed that the processing demands of musical syntax (harmony) and language syntax interact in Broca's area in the left inferior gyrus without leading to music and language main effects. They also observed that a language main effect in Bronca's area only emerged in the complex music harmony condition, suggesting that with our stimuli and tasks a language effect only becomes visible under conditions of increased demands on shared neural resources. In the fundamental linguistic base of music perception the representation of the perceived tones is based on a high level, logic oriented formalism. It also acts as a type of long term memory. Here exists a semantic network of symbols and identification of their relationships is the act of music perception. A hybrid formalism in this context is constituted by two different components:

a. A terminological component describes the concepts like chords, tonic dominants, rhythm, melody, tempo, dynamics etc.
b. An assertional component stores information concerning a specific context.

A number of conceptual spaces operate between conceptual and linguistic areas. Such conceptual spaces function as a workplace in which low-level and high-level processes access and exchange information respectively from bottom to top and top to bottom. In case of music perception, the knoxel in the music conceptual space changes its position when the perceived sound changes pitch or timbre. Consideration of the partials of a tone allows us to deal with microtonal tones, embellished notes, rough notes etc.

4. Perception Based Cognition and Music Perception:

When we listen to a musical composition our auditory perceptions are translated into signals and transmitted to the brain. These signals are further translated into abstract, code-like, non-representational 'language of the brain'. These are then processed by rule governed algorithms based on formal logic. Barsalou views that these code-like, algorithm-based (amodal) systems have appealing features (1999a). They can represent types and tokens, produce categorical inferences and represent propositions and abstract concepts. Barsalou also advocates for an explicitly embodied cognition. In this view, the perception based cognition begins with the biological brain's interactions with the world via sensation and physical action. A parallel neural system performs thinking. This neural system simulates perception down to the level of raw sensory and motor interactions with the world (Barsalou 1999a). Hence music cognition depends on *'perceptual symbols'* that are activated neural groups distributed throughout the sensory motor areas of the brain. At the basic level musical beats may represent raw sensory experiences like visual, audial, tactile, motor control etc.

Much like visual perceptions, in our auditory perceptual modality also diverse elements like shapes and motions are separately processed. They are integrated in more advanced phase of cognitive processing in associative areas or *'Convergence Zones'*. Here they are further

processed and form a single *'experienced reality'*. All these are very much controlled by *'proprioceptive experiences'* or current body states. Barsalou (1999a) views that this time three types of 'introspective experiences function simultaneously:

a. Representing an experience in its absence through memory or imagination,
b. Cognitive activities such as rehearsal, retrieval, elaboration, comparison etc,
c. Emotional states or moods.

Combination of *'proprioceptive'* and *'introspective'* experiences leads to the development of abstract concepts and varied related emotional arousals. But in case of music perceptions the formation of perceptual symbols is subject to selective attention. Barsalou further observes that perceptual symbols are formed in circuits that parallel the circuits through which primary perceptions are processed. These parallel circuits are capable of reproducing a more or less detailed simulation of actual events as they have been and might be experienced in parallel circuits. So memories of similar experiences are organized around integrated systems of perceptual symbols called *'frames'* and *'schemas'*. Limitless simulations can be developed by these schemas and their associated perceptual symbols. It leads to a reconstruction of experience that is 'parabolic' in nature. Mark Turner's idea of *"parabolic projection"* (1996) exposes how readers use their everyday experiences to reach an interpretation of a text world. Subsequently it helps the reader/listener to project that interpretation into real life situations. In this way, conceptual projection of a story and narrative in parabolic fashion is a fundamental building block to our understanding of the world as well as the way they operate as human thinking and feeling. Within any artistic creation like a literary piece, a portrait or even in a musical composition *"parabolic projection"* operates intertextually. Thus this unconscious employment of parable is a fundamental and continuous cognitive instrument of thought processing not only in understanding an artistic creation but also in real life situation. If this idea is applied to the functioning mechanism in music perception, it can be found that embedded stories in music perception are a kind of *"metaphor in narrative form"* that challenges the ingrained perspective of the listeners. At the same time, it projects extra domains of knowledge into the existing world views of the listeners and results in a modification in the interpretative cognitive models of the listeners (Burke, Michael, 2011, PP-116). Here the central grounding is the mechanism of the cognitive projection of short spatial stories. Projection of one short mental space helps listeners to understand other mental spaces or even to create a new mental space. It also occurs in our everyday cognitive acts in real life situations. When two or more mental spaces combine or interact, there is a *'conceptual blending'*. Such fundamental cognitive understanding is massively controlled by *'schemas'* and *'frames'* and entail many other patterning like prediction, evaluation, planning etc. Such *'frames'* and *'schemas'* also dominate the cognitive context. The forging subject matter of short term memory is also controlled by them. Thus schemata play vital role in the determination process of which elements of a perception will be attended to as well as which elements of a concept will be activated by a perceptual simulator. *'Cognitive frames'* are culture-specific activation within members of a group or culture. A particular schema might build the foundation of a particular frame. The activated frames in return will control how a particular musical pitch or tempo might be interpreted. In some events, a particular message or group of messages may totally transform the 'frame'. Coulson (2011) calls this *'frame shifting'*.

5. Basic Cognitive functioning in Music Perception:

Some fundamental cognitive processing allows us to organize and understand the world. Recent researches on them lay emphasis on three cognitive aspects namely *'categorization', 'cross-domain mapping'* and *'conceptual models'*. These aspects can provide an elemental framework operating in the cognitive function of music perception. In our everyday real life situation a basic cognitive process is our ability to categorize things. Shapes, sounds, smells, tastes, sensations, movements etc are consciously or unconsciously categorized at every moment. In a word, categorization occurs in all sensory modalities and throughout the range of mental activities. Likewise, in case of music perception, conceptualization begins with the task of categorization. Hence Zbikowski rightly observes that musical categories are concepts (2017, PP- 60). The process not only involves categorization but also an additional ability to identify the relationality between categories. Here one vital idea is that the process of categorization is not a rigid formation. They may not simply reflect the structure of global knowledge formation because the formation of such categories is a result of our interaction with the surrounding and hence they are constantly redefined and modified. Moreover, there is no sharp boundary between two such categories. In case of music perception, the weighted and unweighted collection of the most typical attribute-values forms the prototype. For example, the collection of beat, meter, tempo, syncopation etc forms the prototype of 'rhythm'. The collection of crescendo, decrescendo etc forms the prototype of *'dynamics'*. On the other hand, *'melody'* is constituted with the typical collection of pitch, theme, conjunct, disjunct etc. Similarly, the collection of smaller categories like chord, progression, consonance, dissonance, key, tonality, atonality etc forms the prototype of *'harmony'*. Another prototype *'tone color'* is formed by register, range, instrumentation. *'Texture'* may be monophonic, homophonic, polyphonic and constituted by imitation, counterpoint etc. Binary, ternary, strophic, through-composed etc shape the elemental prototype of *'form'*. As Zbikowski (2017) observes, categories are organized into hierarchal taxonomies where the *'basic level'* is neither the lowest nor the highest level of a taxonomy but rather an in-between level of maximum utility. During the categorization process, some other cognitive abilities like aggregation, description, membership prediction etc keep functioning.

Another fundamental cognitive processing i.e. *'cross-domain mapping'* is a process of structuring of our understanding of unfamiliar domains with the help of familiar domains. For example, by comparing the speed of the beats in a musical composition we can understand the *'tempo'*. Simultaneously, by comparing the relative loudness or quietness in a music composition we can get idea of the *'dynamics'*. Even when we map down the pitch, we can also understand the coherent relationship between small physical spaces within the musical composition. Here the most vital role is played by *"image schemata"*. Proprioceptive experiences work as vital grounds for these *'image schemata'*. In every image schematic structure a hidden source domain and target domain is operational.

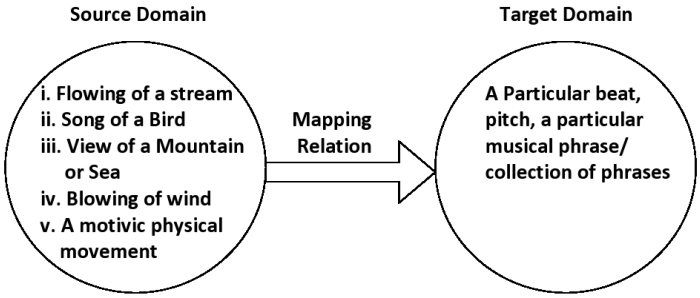


Figure- 2; Mapping Relation between source and target domain

The more image schematic structures are there, the better cross-domain mappings will be there. Inevitably, a large depository of socio-cultural framework operates within such embedded domains. One thing must be remembered here that like other mapping relations, such cross-domain mappings are not absolute; rather they reflect partial alignment. Despite such limitations, cross-domain mapping co-relates musical concepts with non musical domains. It also develops the ground for conceptualizing elusive musical phenomenon.

To deal with the cognitive process of *'conceptual blending'* I must invoke Fauconnier and Turner who proposed that the dynamic process of 'conceptual blending' involves small interconnected conceptual pockets known as *'mental spaces'* (2002, PP-39ff). Such small *'conceptual pockets'* are connected to *'long term schematic knowledge'* known as *'frames'* as well as to *'long term specific knowledge'*. There are at least four *'mental spaces'* that include two *'input spaces'*, a *'generic space'* and a *'blended space'*. The *'generic space'* contains the common elements of two *'input spaces'*. The *'blended space'* contains some elements from each input space.

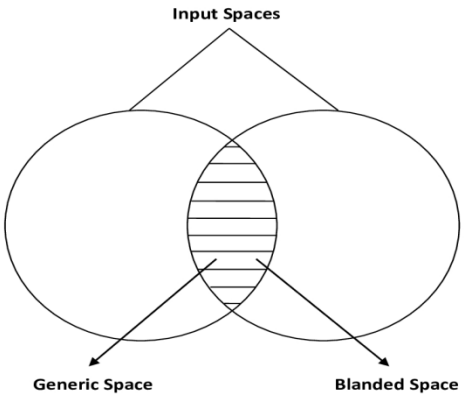


Figure- 3; Mental Spaces

The *'blended space'* may also contain *'emergent structure'* i.e. additional elements that can include new elements retrieved from long term memory or resulting from comparison of elements drawn from the separate inputs (Ritchie, L. David, 2004a). In case of visual perceptions of complex objects like a hammer, their mutual positions and shapes are important in order to describe perceived object.

Figure- 4; Visual Perception of hammer

In case of music perception, the mutual relationships between pitches and the timbres of the perceived tones are important in order to describe the perceived chords. Thus spatial relationships in static scenes analysis are in some sense analogues to sound-relationships in conceptual space of music. *'Conceptual blending'* gradually builds conceptual models. The cognitive process initiated by *'categorization'* and elaborated by *'cross-domain mapping'* now gradually leads to the development of such models. It also develops the basement for reasonable conceptualization. In Zbikowski's opinion, the resultant conceptual blend possesses structural features that may not be part of either of the input spaces due to the operations of *'composition, 'completion', elaboration'* etc. (Zbikowski 2002, PP-80-81). When 'composition' occurs, aspects of the input spaces join hands to form new entries. During the task of 'completion' background knowledge fill in gaps and extend the blended space. The structure of the *'blended space'* is developed further in the task of 'elaboration'. This time the input spaces lose their vitality and a more autonomous entity comes into existence. All these activities operate within a specific cognitive context that includes socio-cultural-biological-physical aspects of our existence. Such higher cognitive organization takes the attributes like rhythm, dynamics, melody, harmony, texture, form etc to a more abstract realization and creates aesthetic sensibilities.

6. Functioning of the Musical Text World in Songs and Instrumental Music:

'Text World Theory', mainly proposed by Werth (1999), Gavins (2007), Semino (1997), Whiteley (2011) et al aims at providing a holistic framework for an analysis of the way discourse is conceptually constructed and negotiated by discourse participants (in case of music the composer, producer and the audience). Within this cognitive framework, there is ***"Discourse World"***. Werth defines this as *"the situational context surrounding the speech event itself"* (1999:83). Rather than merely dealing with the location, time, discourse participants etc, it considers the relationships, knowledge, experiences the participants use to understand the semiotic features. There is a ***"Text World"*** which is defined by Werth as *"a deictic space, defined initially by the discourse itself and specifically by the deictic and referential elements in it"*. Within the 'Text World' some features of language that establish the parameters and contexts of the text world are known as *'world building elements'*. At the same time, some events, characters, temporal features etc contribute to progress the world building process. These elements come under the umbrella term of *'function advancing propositions'*. Besides these two worlds, there are also ***'modal worlds'***. Among them, *'Epistemic modal worlds'* negotiate knowledge between the discourse and text world, *'Deontic modal worlds'* are

formation of mental models with varying degrees of obligation and *'Boulomatic modal worlds'* are hypothetical or unrealized *"wish"* worlds. In this section I shall try to view how these worlds are formed in music and how it can shed light on the role of embodiment and culture in the interpretative process.

In the foresection I have discussed Fauconnier and Turner's notion of *'mental spaces'* (2002). Making a further elaboration of *'mental spaces'* I shall try to observe how such spaces are separately derived in a song and in an instrumental music composition from the proposition in the musical text world and how they are established through deictic parameters and referential elements. Here is the lyrical text of Bob Dylan's noted song **"Forever Young":**

May God bless and keep you always
May your wishes all come true
May you always do for others
And let others do for you.
May you build a ladder to the stars
And climb on every rung
May you stay forever young
Forever young, forever young
May you stay forever young.

May you grow up to be righteous
May you grow up to be true
May you always know the truth
And see the lights surrounding you.
May you always be courageous
Stand upright and be strong
May you stay forever young
Forever young, forever young
May you stay forever young.

May your hands always be busy
May your feet always be swift
May you have a strong foundation
When the winds of changes shift.
May your heart always be joyful
May your song always be sung
May you stay forever young
Forever young, forever young
May you stay forever young.

From the moment a listener starts listening this song, he/she is engaged perspectively in encountering diverse socio-cultural-biological elements. In the listening context of musical composition the listener occupies the *'here and now'* position or the deictic *'origo'* or centre and the performer occupies the *'then'* position of the context of composition. As soon as the listening act starts socio-cultural knowledge, experiences and body-based orientation schemata start functioning. All these are context sensitive and allow incremental transfer of knowledge from private to public domain. The listener tries to build a possible world that is rule-governed

and corresponds to our actual world. Within this formation the current bodily condition of the listener is heavily involved. In Wertth's observation: *"all approaches to cognitive discourse study are founded on the basic assumption that the mind and the body are inextricably linked"* (1999:36). So it is evident that the listener deliberately tries to map his/her physical experience onto unfamiliar situations of *'image mapping'*. Thus the deictic *'origo'*, very often the listener's *'here and now'* condition or more explicitly the listener's positioning of the *'self'* in space and time is the beginning point of the text world mapping.

When the musical discourse world advances and becomes more complex, i.e. the listener is more absorbed in his/her cognitive functioning, he/she witnesses a kind of *"empathetic identification"* with the fictional character(s) within the text world. It is a major source of emotional outbreak within the listener. To deal with such *"experiential significance"* Joanna Gravins speaks of *"text drivenness"* that can account for such activation of a variety of knowledge. The more variations are there in the musical genre, the greater amount of experiential knowledge will be utilized by the listener to deal with the context, expectation, cultural and linguistic constraints etc. The listener now tries to identify *'the world building elements'* based on the relationships projected by the embodiment of the *'origo'* or centre ('here' and 'now' conditions). At the same time , the *'function advancing propositions'* involving the characters in the song, the narrator, actions depicted in the song, expected illocution in the speech act, temporal shifts etc will also be evaluated by the listener. Thus the text world may be considered to be a product of the discourse world. During all these activities the concept of *"modality"* continues its functioning. This *'modality'* is directly related to speech act theory particularly to Gricean maxims and the desired illocution by the composer. If the narrative adopts a different temporal or spatial perspective or if there is a deictic triggering mainly by the use of temporal or spatial adverbs like yesterday, today, tomorrow, forever, then, now, here, there etc, the modal sub-world may witness a switch. At a higher degree such exploitation of socio-cultural experiences, embodiment and identification of motives etc will lead to the formation of unidirectional hermeneutical influences or ideological content by the listener. This time the ontological division between the *'text world'* and the *'discourse world'* is negotiated to such an extent that the boundary between them looks blurred. This kind of perspective taking by the listener makes significant transformation in the understanding of everyday life. This hovering within the text world often serves as a catalyst and enables listeners to connect the fictional world to own personal experiences. Even awareness of emotional outbreaks has a vital therapeutic application. Emotional awareness does not merely involve thinking about a particular feeling. Rather it involves a conscious act of observing the particular feeling. It may give a better access to suppressed or inhibited experiences and promote assimilation in self-conscious experiencing in schematic frames. This potential act is a basic act in music therapy that often aims at a particular moulding in future actions and retrospective and introspective thinking.

But the cognitive functioning in instrumental music perception is a bit different because it is massively dependent on intuitive cognition. Both instrumental music and language are syntactic systems that employ complex and hierarchically structured sequences that are built by using implicit structural norms. In such act of music perception, there is a continual stream of perceptions and actions. These include introspective simulators like stream of thoughts and perceptual flow of sound. Though there is a continual stream of perceptual simulators, not all of them interact with the stream of perception and action. Here are also some maintenance

operations like topic management, turn taking, keeping track of levels (layering) etc. Such operations move in and out of conscious attention of the listener and interact with the perceptual simulations of *"stream of consciousness"*. This time schematic information that is highly salient is not always activated. Schematic information that is of little importance or not immediately relevant may also be partially or totally activated. It is an established fact in neuro-scientific studies that implicit learning processes are the cognitive substrate of social intuition. Though intuitions may result in sub-optimal decisions, in case of instrumental music perception it has a vital role because instrumental music perception involves exploitation of subjective experiences or schematic frames that are mainly associated with knowledge or experiences gained through implicit learning. The conceptual correspondence between implicit learning and intuition is further established during perception of instrumental music such as musical compositions in flute, sarod, santoor, sarengi, sitar, violin and so on because it demands consideration of the unconsciously retrieved schematic frames. The cognitive processing ultimately leads to a probabilistic structure of the musical cues and resultant aesthetic sensibilities. Unlike songs, here the parabolic structure formation is partial. So the aesthetic and emotive aspects do not have an explicit and conscious structure. When the listener lacks sufficient knowledge regarding tonality, pitch or rhythm of an instrumental music composition, he/she may get lost at a certain musical phrase that has ignited considerable amount of schematic frames within the listener. In other words, the listener may get lost in complex webs of environmentally stored information both in mental and public representation. Another option may be that the listener is consciously following the musical beats, pitches, rhythm etc of every musical phrase. Here the activation of schematic frames will be partial and short-termed. The listener will not get a logical sequence of his/her subjective experience. Rather, partially ignited schematic frames will generate probabilistic judgement and mystic emotional outbreaks that are completely dependent on those partially ignited schematic frames. Such frames are partially ignited because a certain musical composition (here of course instrumental) has a constant temporal flow. Hence a certain musical syntax may catch greater attention of the listener and may involve him/her to process that phrase in neural net with greater depository of knowledge or schematic frames. Another probability is that the listener is following the total temporal sequence of the musical composition without paying much attention to a particular syntax. This time the ignited schematic frames will not have sufficient scopes for a detailed processing. Here the non-verbal decoding skill is massively controlled by intuitive processing of phenomenological awareness. There will be an inaccurate inference drawing and it will vary from person to person, time to time and from culture to culture depending on the cognitive environment of the listener. After a few such incremental processing, it will gain habitual patterns triggered automatically by situational musical cues. Thus the perception skill here is basically intuitive that involves temporal sequencing and prediction that may ultimately lead to a condition of *"emotion as information"*.

7. Conclusion:

Right from infancy human cognition is highly interactive with both the physical and social environment. In case of non-musicians, the cognitive functions from the initiation of the listening act to the gaining of aesthetic sensibilities are sequential interactions of bodily capabilities (including genetically determined brain capabilities) with the external environment. Moreover, connections in sensory-motor systems are greater in number than the system that integrates conscious subjective experiences. For this reason, recruitment-learning in music

perception flows from sensory-motor to subjective experiences, not the reverse. In other words, the nonphysical is conceptualized in terms of the physical. Complex networks of perceptual simulations in music perception not only become activated but also activate other cognitive tasks like community based extended cognition and formulation of a particular cultural hermeneutics when it involves larger socio-cultural structures. It is a continuous evolutionary process of human brain that uses the external environment interactively than merely making an internal representation. So perception of music by non-musicians is also an evolutionary process represented by social forms of extended cognition like teamwork, collaboration, empathy formation etc. This also works as a motivation to go into such cognitive exercises frequently and creates a taste for music in the non-musicians. Subsequently it becomes an integral part of social context. The limitation of this study is that it can not explain how neural synaptic links are formed among the group neurons that constitute perceptual simulators.

8. References:

Abrams, D.A., Ryali, S., Chen, T., Chordila, P., Khouzam, A., Levitin, D.J., Menon, V. (2013) "Intersubjective Synchronization of Brain Responses During Natural Music Listening", *The European Journal of Neuroscience,* 37(9): 1458-1469.

Barsalou, L. (1999a), 'Perceptual Symbol Systems', *Behavioral and Brain Sciences,* 22: 577-609.

Batsalou, L. (1999b), 'Author's Response: Perceptions of Perceptual Symbols', *Behavioral and Brain Sciences,* 22: 637-660.

Burke, Michael. (2011), *Literary Reading, Cognition and Emotion: An Exploration of the Oceanic Mind,* New York, NY: Routledge.

Coulson, S. (2001), *Semantic Leaps: Frame-shifting and Conceptual Blending in Meaning Construction,* Cambridge: Cambridge University Press.

Gavins, J. (2007). *Text World Theory: An Introduction,* Edinburgh: Edinburgh University Press.

Fauconnier, G. and Turner, (2002), M. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities,* New York, NY: Basic Books.

Kunert, R., Willems RM, Casasanto D., Patel AD, Hagoort P. (2015), 'Music and Language Syntax Interact in Broca's Area: An fMRI Study'. PLOS ONE, 10(11): e141069. https://doi.org/10.1371/journal.pone.0141069.

Ritchie, L. David. (2004), 'Lost In Conceptual Space: Metaphors of Conceptual Integration' *Metaphor and Symbol,* 19: 31-50.

Semino, E. (1997), *Language and World Creation in Poems and Other Texts,* London: Longman.

Werth, P. (1999), *Text Worlds: Representing Conceptual Space in Discourse,* Harlow: Longman.

Whitely, S. (2011), Text World Theory, Real Readers and Emotional Responses to the Remains of the Day. *Language and Literature.* 20 (1), PP- 23-41. (Google Scholar).

Zbikowski, Lawrence M. (2002), *Conceptualizing Music: Cognitive Structure, Theory and Analysis,* Oxford: Oxford University Press.

Zatore, Robert and Peretz, Isabella, (2003), *'The Cognitive Neuroscience of Music'*, Oxford: Oxford University Press.

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

FRACTAL BASED CATEGORIZATION OF BENGALI PHONEMES: A PILOT STUDY

Pijush Kanti Gayen, Shankha Sanyal and Samir Karmakar
School of Languages and Linguistics,
Jadavpur University, India

ARTICLE INFO

Article history:
Received 05/12/2020
Accepted 18/12/2020

Keywords:
Acoustics,
Categorization,
Phonemes,
DFA,
Nonlinear analysis,

ABSTRACT

The natural human speech creates turbulence in the vocal tract during its production. For this reason, the natural speech sounds are thought to be absolutely nonlinear in nature. In this paper, we describe the complexity of nonlinear speech signals using the concept of fractal dimension. We use a multi-scale fractal dimension to quantify the degree of self-similarity of the speech signal. The main attempt here is to categorize the basic phonological oppositions by fractal technique in the Bengali language. Detrended Fluctuation Analysis (DFA) is used here to quantify the long-range temporal correlations present in different chosen speech units of the Bengali language. It is implemented on the Bangla consonants to distinguish the voiced segments [+voiced] and the aspirated segments [+aspirated] from their respective unvoiced [-voiced] and unaspirated [-aspirated] counterparts. The preliminary results indicate that the phonological oppositions scale differently in the acoustic domain. The effective categorization of the oppositions in terms of the respective fractal dimension will result in a better resource compatible with the models of speech recognition in the field of acoustic categorization.

1. Introduction:

The analysis of speech signals with nonlinear methods has been found to be more effective due to the turbulent airflow in the vocal tract, followed by the nonlinear neuro-muscular process in the larynx and other different parts that involve in speech production. A speech recognition system must map the acoustical features of the speech signal into linguistic entities (F. Martinez et al. 2003). A phoneme is generally defined as the smallest contrastive linguistic unit which may

ছন্দ 2020 *Special Issue from selected papers of* ***International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives***

Corresponding Author: Pijush Kanti Gayen
Email: pijushgayen1991@gmail.com

bring about a change of meaning, while a phone is a unit of speech sound that may refer to any speech sound or gesture without regard to its place in the phonology of a language (Coxhead, 2006). Thus, a phoneme is a set of phones or a set of sound features that are thought of as the same element within the phonology of a particular language. Hence, an automated approach towards the categorization of phones in terms of different phonological oppositions of which they are part is very important as this may lead us to a robust speech model corresponding to a particular language or even a group of languages. Accordingly, we need to quantify the value of the different phonological signals to develop a speech model.

In the pronunciation of the Bengali language, aspiration holds a very distinct place. In the famous 'Bhasha Prakash Bangala Vyakaran' (Chatterji 1996, p. 42) it is shown that every second and fourth consonant in the five-strong Bangla 'barga' (a group of five serial sounds in an alphabetic list) are aspirated while others remain unaspirated. Dr Chatterji also classified Bangla consonants on the basis of manner and place of articulation as well as aspiration and voicing. In (Shahidullah 2003, p. 30-31), an attempt was made to classify the Bangla consonants on the basis of aspiration, as it plays a significant role in the subtleties of pronunciation. In (Islam 2002, p. 44) Bangla consonants were classified in terms of aspiration along with other factors like manner and place of articulation, voicing, nasalization and muscular tension. In Bangla classificatory matrix, aspiration and voicing are generally given equal weightage as these two are considered as distinctive features in the Bengali language.

Before proceeding to the main objective of the paper, let us define in detail what is meant by aspiration. In layman's language, aspiration is nothing but the pronunciation of sounds with a little puff of air from the lungs. It is the burst of air that accompanies the release or closure of some obstruent (i.e., closure of pulmonary airflow leading to friction like noise). In (Crystal 2003, p. 37) aspiration is defined as "A term in phonetics for the audible breath which may accompany a sound's articulation, as when certain types of plosive consonant are released." The sounds that have got special phonetic characteristics as such are traditionally called *aspirates*.

The diacritic for aspiration in the International Phonetic Alphabet is a superscript 'h', [^{h}]. In a normal case, the superscript follows the main phoneme. This is, in a stricter sense, post-aspiration. Pronunciation may be characterized by what might be called pre-aspiration when the symbol [^{h}] precedes the main phoneme. Aspiration is considered to be a distinctive feature in Bangla as its presence or absence creates a difference in meaning in otherwise similar-sounding words, so /b/ and /b^{h} / will be considered two different phonemes in Bangla.

For efficient speech recognition processes to be developed in a particular language, it is essential to start from the root, i.e. phoneme classification. In the literature, a number of studies can be found, which use a variety of parameters such as Wavelet-based features (Long and Datta, 1996; Biswas et al, 2015), Support Vector Machines (Yousafzai, 2010), Convolutional Neural Networks (Palaz Collobert & Doss, 2013) and several other ensemble methods (Dekel, Keshet & Singer, 2004; Waterhouse & Cook, 1997), to achieve their goal. Amongst these, Mel Frequency

Cepstral Coefficients (MFCCs) (Davis & Mermelstein, 1980) is the most widely used feature extraction technique. MFCC features represent the spectral shapes of input signals and shows good results in clean condition but the performance of MFCC deteriorates in the acoustic/sensor mismatch condition. To make features more robust in the complex auditory environment, researchers have studied and developed different features such as Perceptually Linear Prediction (PLP) (Hermansky & Morgan, 1994) analysis, Gammatone Frequency Cepstral Coefficients (GFCCs) (Shao et al, 2010). All these works use several linear features or a mix of linear-nonlinear, while no reported work has deployed nonlinear features for speech classification and recognition using acoustic waveforms. In this work, the main aim is to develop a robust methodology using nonlinear speech acoustic features which will lead to effective phoneme identification and classification.

The different sound units lie on an imaginary Cartesian plane. The scaling exponent derived from Detrended Fluctuation Analysis (Peng et al, 1994) is a measure of the time-varying long-range temporal correlations present in the sound units which are phonological oppositions by nature. An offshoot of the method is the fractal dimension which actually is a geometric invariant to predict the relative distance between the sound units on the Cartesian plane (Katz, 1988). Among them, the vowels can be distinguished more easily by the formant analysis. But the difficulty occurs in the case of consonants, as they are just a burst or a closure for a moment. The pitch contour also is not so significant. So the task becomes more challenging when it comes to the categorization of the consonants. Here we attempt to categorize the basic phonological oppositions by fractal technique in the Bengali language. It is implemented on the Bangla consonants to distinguish the voiced segments [+voiced] and the aspirated segments [+aspirated] from their respective unvoiced [-voiced] and unaspirated [-aspirated] counterparts. In this paper, we will discuss the categorization of consonants according to their phonological oppositions (restricted to the voiced-unvoiced and aspirated-unaspirated) using the latest state of the art nonlinear fractal methods.

2. Experimental Data:

We have used the following sets of Bangla consonants along the dimensions of [±voiced] and [±aspirated]:

Sl.	**[-voiced]**	**[+voiced]**
a.	/p/	/b/
b.	/t/	/d/
c.	/k/	/g/

Sl.	[-aspirated]	[+aspirated]
a.	/p/	/p^h/
b.	/k/	/k^h/

We recorded the data according to our framework and pitch profile analysis was performed in the *Praat* software tool. The acoustic waveform (**Fig.1-6**) corresponding to each of the phones was extracted and the Hurst Exponent was evaluated for each of them. The Hurst Exponent is a robust tool with which the inherent self-similar patterns of human speech signals can be quantified using a specific integer value.

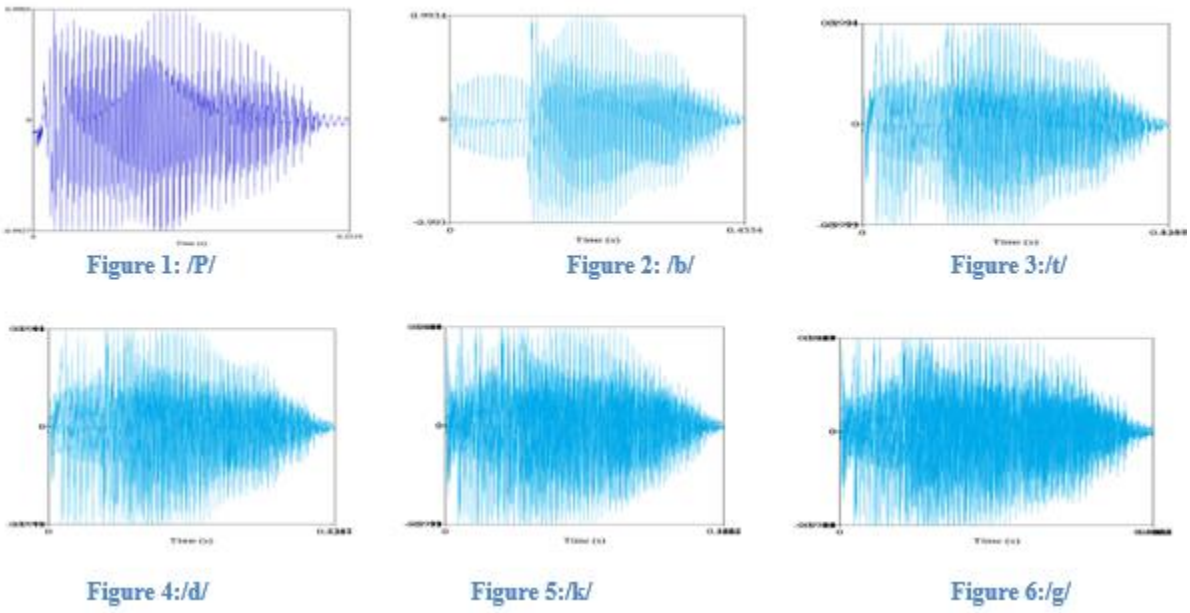


Fig.1-6: Acoustic Waveforms of different voiced and unvoiced Bengali phonemes

3. Methodology:

In this paper, the Hurst Exponent was evaluated using the Detrended Fluctuation Analysis (DFA) technique proposed by Peng et.al (1994). DFA technique was applied following the NBT algorithm used in Hardstone et.al (2012).

$$X(k) = \sum_{i=1}^{k} [x(i) - \bar{x}]$$

Next, X(k) is divided into time windows of length 'n' samples each, and a local least square straight-line fit (**the local trend**) is calculated by minimizing the squared errors within

each time window. Let Xn(k) indicate the resulting piecewise sequence of straight-line fits. Then, the root-mean-square deviation from the trend, the **fluctuation**, is calculated:

$$F(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[X(k) - X_n(k)]^2}$$

This computation is repeated over all possible interval lengths. In practice the minimum length is ranged from 10 samples to a half-length of input data giving two adjacent intervals. The interval length $n = 2^k$ for $k = 4, 5\ldots \log_2(L) - 1$ was set because the power of two based length to input EEG data were used in this experiment. The relationship between the detrended series and interval lengths can be expressed as

$$F(n)\ \alpha\ n^{\alpha}$$

It can be converted into the Hurst exponent $H = \alpha - 1$ and the estimated FD accordingly as $D_{DFA} = 3 - \alpha$ where α is expressed as the slope of a double logarithmic plot $\log_2[F(n)]$ versus $\log_2(n)$

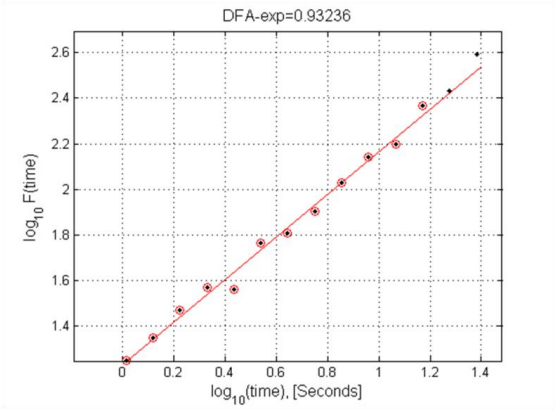


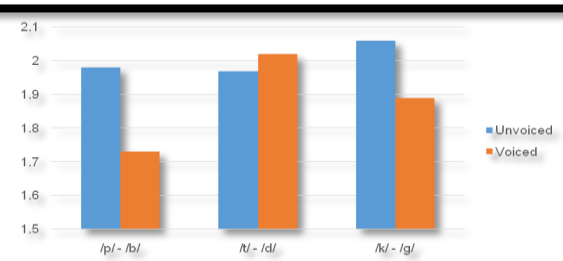
Fig. 7: A double logarithmic plot to evaluate the DFA scaling exponent of audio signals

The scaling exponent provides a quantitative measure of long-range temporal correlation (LRTC) that exists in the audio signals. When the auditory waveform is completely uncorrelated (Gaussian or non-Gaussian probability distribution), the calculation of the scaling exponent results in 0.5, also called white noise. When computing the scaling of the signal profile, the resulting scaling exponent, α, is an estimation of *H*. If α is between 0 and 1, then *x* was produced by a stationary process which can be modelled as a process with $H = \alpha$. If α is between 1 and 2 then *x* was produced by a non-stationary process, and $H = \alpha - 1$. H was evaluated for all sets of phonological oppositions to develop an automated categorization method.

4. Results and Discussion:

From the preliminary analysis, it is seen that each of the phonological opposition sets provides a unique range of scaling exponent values. This indicates that the scaling pattern of each phonological set is different from the other in its speech production mechanism. The development of fractal dimension as a robust mathematical parameter with which different sound units can be categorized is envisaged in this work. This paper reports preliminary findings related to this experiment. The Hurst exponent values corresponding to each phone set are given below (**Fig.8**):

Sl.	Unvoiced	voiced
a.	/p/ =1.98	/b/ = 1.73
b.	/t/ = 1.97	/d/ = 2.02
c.	/k/ = 2.06	/g/ =1.89



Sl.	Unaspirated	aspirated
a.	/p/ = 1.98	/pʰ/ = 1.93
b.	/k/ = 2.03	/kʰ/= 2.12

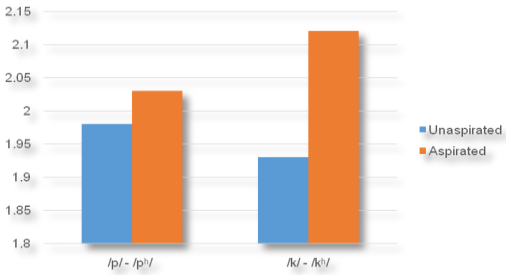


Fig. 8: Hurst Exponent table and graph of phonological oppositions in Bengali

From the figure and associated Table, it is seen that in the case of voiced-unvoiced sound units the categorization is much more significant as compared to aspirated-unaspirated sound units. While the scaling exponent of unvoiced phonemes is on the higher side, the aspirated-unaspirated set gives contrast results, whereas the unaspirated set gives higher complexity values as compared to the aspirated dataset. An interesting observation is that each set possesses a unique value different from the other set. This may enable us to form a specific range of scaling exponent in future with a greater number of pairs, which will help in effective categorization. Analysis with a greater number of variables and a larger recording set would lead to a more conclusive result with a statistically confident outcome.

5. Results and Discussion:

The pilot study reports preliminary findings of fractal-based categorization of Bengali phonemes in terms of phonological oppositions. The initial findings show that the unvoiced aspirated/unaspirated phoneme pairs form a cluster while the same is true for voiced phonemes

both in aspirated/unaspirated scenario. The primary reports of the project are extremely interesting and further analysis with greater number of variables and a larger recording set is being done which would lead to a more conclusive result with statistically confident outcome.

Acknowledgement:
Authors SS and SK acknowledge the DST Cognitive Science Research Initiative (DST-CSRI), Govt. of India for funding the Major Research Project to pursue this Research Work (DST/CSRI2018/78(G))

References:

Barman, D. (2008). The distinctiveness of aspiration in Bangla.

Biswas, A., Sahu, P. K., Bhowmick, A., & Chandra, M. (2015). Hindi phoneme classification using Wiener filtered wavelet packet decomposed periodic and aperiodic acoustic feature. *Computers & Electrical Engineering*, *42*, 12-22.

Chatterji, Sunitikumar. (1988). Bhasha-Prakash Bangala Vyakaran. Kolkata: Rupa and Compan

Coxhead, Peter. "Natural language processing & applications-phones and phonemes." (2006).

Davis SB, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process ASSP 1980;28:357–66. A. Biswas et al. / Computers and Electrical Engineering 42 (2015) 12–22 21

Dekel, O., Keshet, J., & Singer, Y. (2004, June). An online algorithm for hierarchical phoneme classification. In *International workshop on machine learning for multimodal interaction* (pp. 146-158). Springer, Berlin, Heidelberg.

Hardstone, R., Poil, S. S., Schiavone, G., Jansen, R., Nikulin, V. V., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2012). Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Frontiers in physiology*, *3*, 450.

Hermansky H, Morgan N. RASTA processing of speech. IEEE Trans Speech Audio Process 1994;2(4):578–89.

Islam, R. (2002). Bhasha Tatta. Dhaka: Shikha Prakashani.

Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in biology and medicine*, *18*(3), 145-156.

Long, C. J., & Datta, S. (1996, October). Wavelet based feature extraction for phoneme recognition. In *Proceeding of fourth international conference on spoken language processing. ICSLP'96* (Vol. 1, pp. 264-267). IEEE.

Martinez, F., Guillamon, A., & Martinez, J. J. (2003). Vowel and consonant characterization using fractal dimension in natural speech. In *ISCA Tutorial and Research Workshop on Non-Linear Speech Processing*.

Palaz, D., Collobert, R., & Doss, M. M. (2013). Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks. *arXiv preprint arXiv:1304.1018*.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical review e*, *49*(2), 1685.

Shahidullah, Muhammad. (2000). Bangala Bayakaron. Dhaka: Mowla Brothers.